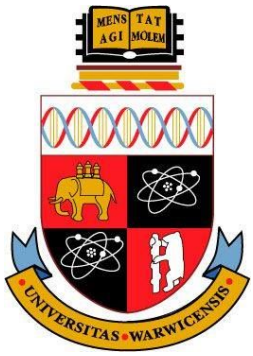
REPLISOME-MEDIATED HOMEOSTASIS OF DNA/RNA HYBRIDS IN EUKARYOTIC GENOMES IS CRITICAL FOR CELL FATES AND CHROMATIN STABILITY

BY **ROWIN APPANAH**

A THESIS SUBMITTED TO
THE UNIVERSITY OF WARWICK
IN PARTIAL FULFILMENT FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY IN
INTERDISCIPLINARY BIOMEDICAL RESEARCH



UNIVERSITY OF WARWICK
WARWICK MEDICAL SCHOOL

OCTOBER 2017

TABLE OF FIGURES

TABLES OF TABLES

ACKNOWLEDGEMENT

‘Facts are facts and will not disappear on account of your likes.’
Jawaharlal Nehru.

‘Most experiments do not work.’
Frederick Sanger

I would like to thank Giacomo De Piccoli for graciously giving me the opportunity to carry out my PhD in his lab and for going above and beyond the duties of a supervisor. Doing so was the catalyst for putting my PhD back on track. I would like to also thank him for his constant intellectual input, kind encouragement and advice that helped to bring this work to completion.

I would like thank Jonathan Millar who was always generous with his time and advice during the PhD. I also thank him and Andrew McAinsh for their crucial advice some one and a half years into the PhD.

I am also grateful to past and present members of the De Piccoli lab for making the lab an enjoyable and convivial place to carry research and for the help they afforded me throughout: Alicja Winczura for her help with Western blotting in the early days, Edward Miller for his tremendous help with cloning, Katie Stokes for help with fission yeast experiments, Nicholas Sillett for restocking shelves with medium and countless little favours as well as Emma Lones for sharing some of her preliminary results with me. Kind thanks also goes to Jürgen Zech for his help in the early days.

I would like to thank the University of Warwick for generously funding my PhD through the Chancellor’s International Scholarship scheme and the department of Warwick Medical School for funding the fourth year of my PhD. I would also like to acknowledge John Davey and Jacob Dalgaard both of whom were instrumental in me joining the University of Warwick in the first

place. I would like to thank the rest of the student and staff body in both the School of Life Sciences and Warwick Medical School for helping to run the departments smoothly.

I would also like to thank my parents and my sister whose love and unconditional support have been much appreciated over the course of these four years (and long before). I also credit them for reminding me that it is OK to step out of the lab from time to time.

DECLARATION

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

The experiments described in Figures **5.8** and **5.9** have been carried out by Emma Lones and Giacomo De Piccoli (Section **5.11**). This has been acknowledged as such within the text and the corresponding figure legends.

ABSTRACT

During DNA replication, forks often stall upon encountering obstacles blocking their progression. Cells will act to speedily remove or overcome such barriers, thus allowing complete synthesis of chromosomes. This is the case for R-loops, DNA/RNA hybrids that arise during transcription. One mechanism to remove such R-loops involve DNA/RNA helicases.

Here, I have shown that one such helicase, Sen1, associates with replisome components during S phase in the model organism *S. cerevisiae*. I demonstrate that the N-terminal domain of Sen1 is both sufficient and necessary for the interaction of the protein with the replisome. I also identified Ctf4 as one of at least two replisome interactors of Sen1. By mutational analysis, a mutant of Sen1 (Sen1-3) that cannot interact with the replisome was created. This mutant is healthy on its own but is lethal in the absence of both RNase H1 and H2. Overexpression of the *sen1-3* allele from the constitutive *ACT1* promoter is able to suppress this synthetic lethality, suggesting that Sen1 travels with replisomes in order to be quickly recruited at sites of R-loops that impair fork progression so as to remove those R-loops.

In some cases, cells exploit fork stalling for biologically important processes. This is the case in *Sz. pombe*, where an imprint prevents complete DNA replication, triggering cell-type switching. This imprint is dependent on Pol1, a component of the replisome. Importantly, a single imprinting-defective allele of *pol1* has been identified to date. Using *in vitro* assays, I have shown that this Pol1 mutant has reduced affinity for its substrates and is a correspondingly poor polymerase. By generating novel alleles of *pol1*, I have also demonstrated that switching-deficiency correlates with the affinity of Pol1 for its substrates *in vivo*. Finally, two interactors of Pol1 ($Mcl1^{Ctf4}$ and $Spp1^{Pri1}$) have been shown to have switching defects.

NOTE ON NOMENCLATURE CONVENTIONS

S. cerevisiae and *Sz. pombe* have similar yet distinct genetic nomenclature conventions. Given that both model organisms were used in this study, it is important to highlight the conventions for both organisms to prevent confusion.

In ***S. cerevisiae***, wildtype gene names are expressed as a three letter, uppercase and italic name followed by a number (e.g. *SEN1*). The three letter name often corresponds to the screen through which the gene in question was originally identified. Mutants are generally designated with the same three letter but in lower case (unless the mutant is dominant) and with an allele designation (e.g. *sen1*Δ, *sen1-1* and *sen1-2*). Because of historical context, the allele designations vary in format (e.g. *leu2-3,112* is a mutant of *LEU2*). Protein names are given as a three letter name with the first letter in uppercase (e.g. Sen1). This is also true for mutant proteins, with the added allele designation (e.g Sen1-1 and Sen1-2). In this study, I have generated constructs of the *SEN1* gene and these constructs are referred to as *SEN1 (X*-Y), where X and Y refer to the first and last residues being encoded for. The corresponding proteins are referred to as Sen1 (X-Y). Different promoters have been used and, where appropriate, the promoters are expressed similarly to their wildtype gene names (e.g. *GAL1*, *SEN1* and *ACT1*).

In ***Sz. pombe***, wildtype gene names are expressed as a three letter, lowercase and italic name followed by a number (e.g. *pol1*). Mutants are generally designated in the same format but with an allele designation. Like in *S. cerevisiae*, the allele designation varies widely (e.g. *pol1-1, pol1-H4 and pol1-ts13*). Additionally, because of the historical context, some (but not all) alleles of *pol1* are referred to as *swi7* to reflect the fact that they are defective for cell-type switching. Similar to the situation in *S. cerevisiae*, proteins names are given as a three letter name with the first letter in uppercase for both wildtype and mutants (e.g. Pol1 and Swi7-1). Sometimes, for the sake of comparison, genes or proteins are referred to their *S. cerevisiae* orthologues (e.g. *swi1*^{TOF1} and Swi1^{Tof1}, respectively).

Several protein tags have been used in this study. When written in gene form, they were written in capital letters and italicized, **irrespective of the host** (e.g. *5FLAG*) and when in protein form, they were written in capital, **irrespective of the host** (e.g. 5FLAG).

CHAPTER 1: INTRODUCTION

DNA replication is essential for cell viability. When cells duplicate their chromosomes, the information needs to be fully and faithfully duplicated. To do so, cells must be able to unwind the entirety of DNA molecules, thus overcoming the many proteins bound to the DNA and the secondary structures that single-stranded DNA can assume. Mounting evidence indicates that a major obstacle to fork progression are DNA/RNA hybrids formed during RNA transcription. Understanding how this occurs will be one of the major focus of this thesis. DNA/RNA hybrids, however, are also essential for DNA replication, since DNA synthesis requires an RNA primer. These primers are removed before the end of DNA replication as RNA molecules are more labile than DNA and stretches of RNA might impede the enzymatic activities of processive DNA polymerases. Some organisms, however, have evolved to use RNA nucleotides incorporated into the genome as a marker for programmed DNA recombination. A notable example of this phenomenon occurs at the mating type locus (*mat1*) in *Schizosaccharomyces pombe*. The mechanisms used to establish this marker and the role of DNA Polymerase α in this process is also a major focus of this thesis.

In this introduction, I will briefly introduce the **eukaryotic cell cycle**, before focusing on the S phase and the process of **DNA replication**. I will introduce several components of the replication machinery (**the replisome**) and will then go on to describe several **obstacles to DNA replication** that forks must overcome during DNA synthesis. I will focus particularly on **(1.) programmed replication fork DNA-protein barriers** and **(2.) DNA/RNA hybrids as barriers**. As an example of the former, I will introduce the **mating-type switching mechanism of *Sz. pombe*** and how cells use such a barrier to create an RNA imprint (or marker) that causes DNA recombination in the subsequent cell cycle. As an example of the latter kind of barriers, I will focus on the **obstacle represented by RNA transcription** to fork progression. I will describe how cells cope with R-loops (an intermediate of the transcription process), and focus on the function of the DNA/RNA helicase **Sen1**. Finally, I

will describe the known functions of Sen1 including in transcription termination and in maintenance of genome stability.

1.1 THE CELL CYCLE AND DNA REPLICATION

The life cycle of a cell follows an ordered and pre-programmed pattern of events known as the cell cycle. Cells are fated to follow this sequence of events unless under special circumstances. These include quiescence (such as in terminally-differentiated neurons) and loss of genetic material (such as erythrocytes). The emergence of the cell cycle in evolution is ancient and serves the crucial purpose of quality-controlling newly-duplicated DNA prior to irreversible division of genetic material between two sister cells. Notwithstanding quiescence (G_0), the cell cycle can be divided in interphase and mitosis where DNA is packaged in chromosome pairs that can be conveniently divided. Interphase can be further divided in G_1 (pre-DNA synthesis gap), S phase where cells duplicate their DNA and G_2 (post-DNA synthesis gap). The boundaries between the different phases are not always easily discernible. For instance, whilst cytokinesis coincides with the end of the mitotic cycle in most organisms, in the fission yeast, *Schizosaccharomyces pombe*, cytokinesis is delayed until the end of the subsequent S phase (Fig 1.1).

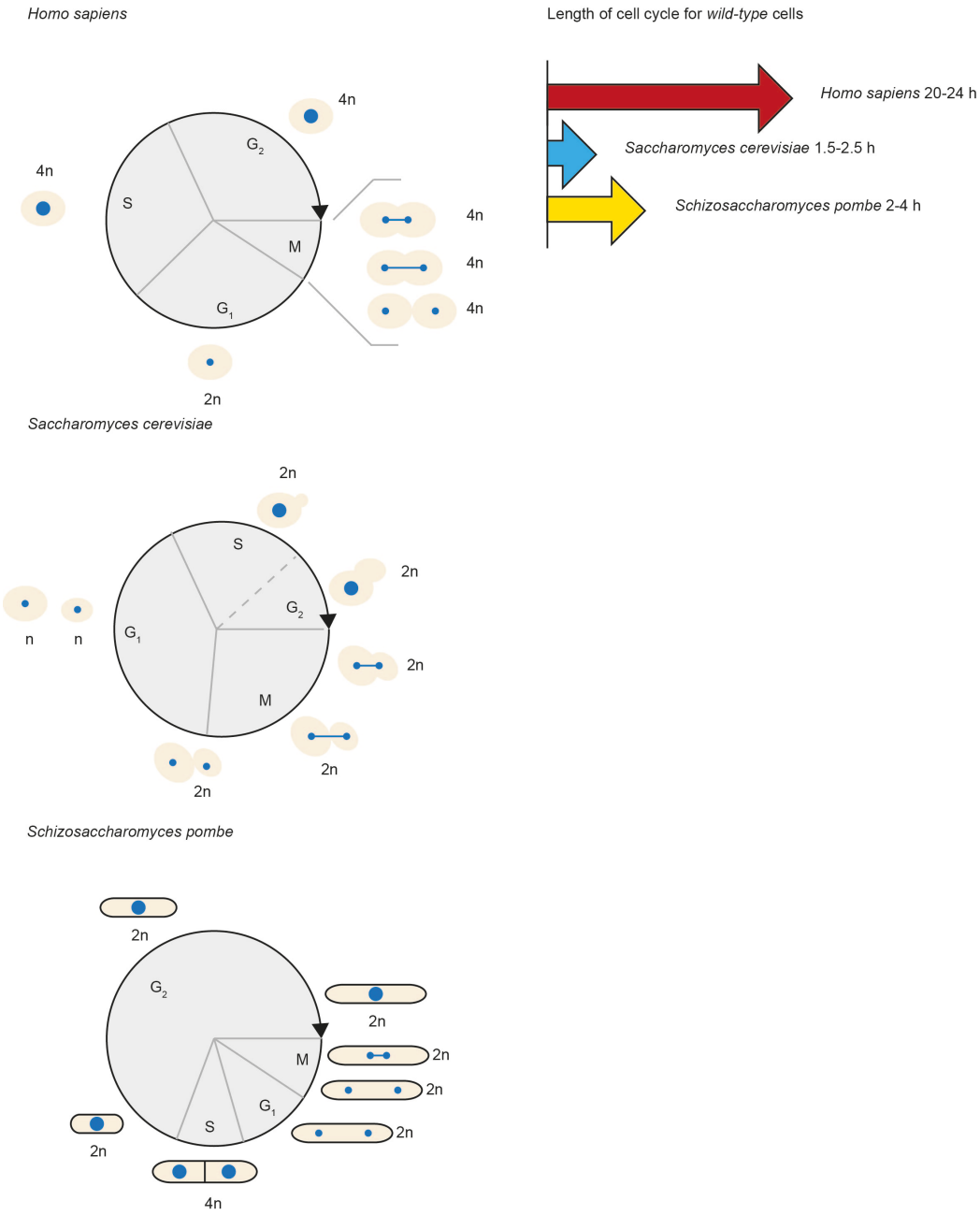


Figure 1.1. The cell cycle follows a patterns of events that is species-specific. The duration of the cell cycle also varies according to the species and, in multicellular organisms, according to cell types. Typically, human cells will complete one cell cycle in 24 h whilst the model organisms *S. cerevisiae* and *Sz. pombe* will go through several cell cycles in a day.

Progression from one phase of the cell cycle to the next is stringently regulated and is mediated through the interplay of activation of cyclin-dependent protein kinases (CDK) and their inactivation by the anaphase-promoting complex/cyclosome (APC/C). The protein kinases are made up of two components: a cyclin-dependent kinase and its cyclin co-factor. Several phase-specific cyclins have been identified. In the model organism *Saccharomyces cerevisiae*, the cyclin-dependent kinase that regulates the cell cycle is Cdc28 and it is activated by nine different cyclins (Clb1-6 and Cln1-3) (Mendenhall and Hodge 1998).

Three of these cyclins (Cln1, Cln2 and Cln3) are required at the G_1-S boundary. Cln1 and Cln2 are strictly G_1-specific and are involved in the decision to commit cells to the mitotic cell cycle, cell growth, degradation of the kinase inhibitor Sic1 and other functions. Unlike both *CLN1* and *CLN2*, *CLN3* is expressed throughout the cell cycle but its transcription is slightly raised at the M-G_1 boundary (Tyers et al., 1993). In fact, Cln3-Cdc28 is required for transcription of *CLN1*, *CLN2*, *CLB5* and *CLB6*. However, Clb5-Cdc28 and Clb6-Cdc28 are inactivated by Sic1 until the latter is degraded by Cln1-Cdc28 and Cln2-Cdc28. This allows Clb5 and Clb6 to initiate S phase. Two other cyclin genes, *CLB3* and *CLB4*, are transcribed optimally at the onset of S phase and plateau until late anaphase. Clb3-Cdc28 and Clb4-Cdc28 seem to contribute partially to the initiation of S phase but are also involved in spindle formation. A final pair of cyclin genes (*CLB1* and *CLB2*) exist whose transcriptions peak prior to anaphase. Both Clb1-Cdc28 and Clb2-Cdc28 contribute, albeit to different extents, to the mitotic event.

Besides cyclin-dependent kinases, the APC/C is another major player in cell cycle progression. The APC/C is an essential, multi-subunit (5 subunits in *Drosophila melanogaster*, 12 subunits in vertebrates and 13 subunits in *S. cerevisiae*) E3 ubiquitin ligase (Peters, 2006). The APC/C influences progression through the cell cycle by polyubiquitinating cyclins, targeting them for degradation by the 26S-proteasome. Importantly, the APC/C requires a co-factor to be functional. Indeed, by metaphase, the APC/C is phosphorylated by CDK, allowing it to associate with one of its co-factors, Cdc20 (Peters,

2006). APC/C^{Cdc20} selectively tags S phase and mitotic cyclins for polyubiquitination. This inactivates the CDK, allowing its substrates to undergo dephosphorylation, a critical event for the progression of mitosis. APC/C^{Cdc20} also mediates the degradation of securin. Securin is an inhibitor of separase, a protease that digests a subunit of cohesin (a complex that holds sister chromatids together). As such, the APC/C is essential for the transition from metaphase to anaphase. Cyclin degradation leads to dephosphorylation of the APC/C. This promotes its dissociation from Cdc20, thus permitting its association with a different co-factor, Cdh1. APC/C^{Cdh1} degrades Cdc20, promoting origin licensing in G_1. Once cells enter S phase however, phosphorylation of APC/C mediates the dissociation of Cdh1 from the cyclosome, greatly reducing its activity in S phase (Peters, 2006).

Successful completion of the cell cycle necessitates the satisfactory completion of individual phases within the cell cycle prior to progression to subsequent phases. For instance, fusing of S and G_2 cells triggers premature condensation and mitotic entry of S phase nuclei, leading to chromosomal fragmentation (Rao and Johnson, 1970). Meanwhile, defects in chromosome attachment to the spindle lead to aneuploidy and stalling of replication forks can lead to un-replicated stretches of DNA, eventually causing non-disjunction between sister chromatids and DNA double-stranded breaks. Thus, to ensure that the different processes within the cell cycle occur in the correct order and with tolerably low levels of error, cells have evolved several monitoring points known collectively as cell cycle checkpoints. Checkpoints involve delaying the entry into a subsequent phase of the cycle until some threshold is reached or until a particular problem is dealt with. For example, cell size checkpoints depend on cells reaching a certain critical size and these are known to exist in G_1 and G_2, but their position within the cycle is species-specific (Barnum and O'Connell, 2014). The intra-S phase checkpoint plays a crucial role at stalled replication forks by regulating and protecting them, and preventing the progression into the G_2/M phases of the cell cycle. The DNA damage checkpoint, similarly, stalls entry into the following phase of the cell cycle until damaged DNA is repaired. Finally, the mitotic spindle checkpoint prevents premature degradation of cyclins and securin via the APC/C^{Cdc20} upon

unproductive occupation of kinetochores in metaphase. The different checkpoints are summarized in (Barnum and O'Connell, 2014). Whilst the discovery of cyclins was made in sea urchin eggs by Timonthy Hunt (Evans et al., 1983), the CDK was originally identified and cloned in *Sz. pombe* by Paul Nurse (Nurse and Bissett, 1981) and mutants of the genes required for cell cycle control (*cdc*) were originally identified in *S. cerevisiae* by Leyland Hartwell (Hartwell et al., 1973). Hunt, Nurse and Hartwell would go on to share the 2001 Nobel Prize in Physiology or Medicine for their discoveries of key regulators of the cell cycle.

Interestingly, the S phase represents the most dangerous part of the cell cycle. During replication, DNA is unwound by means of the replicative helicase and single-stranded DNA (ssDNA) is stabilized by recruitment of the replication protein A (RPA). This prevents re-annealing prior to DNA duplication. However, in this configuration, the DNA is especially prone to damaging agents. Indeed, in yeast, a marker for DNA damage, Rad52-foci, is elevated several fold in S phase compared to other parts of the cell cycle (Lisby et al., 2001). Nonetheless, DNA replication is an absolute necessity as it duplicates genetic material to be shared precisely equally between two sister cells.

In the following chapter, I will illustrate how eukaryotic cells ensure the complete and faithful duplication of their chromosomes.

1.2 DNA REPLICATION

1.2.1 ORIGIN FIRING

Prior to entering S phase, origins of replication (loci within the genome from which replication is initiated) need to be licensed. In eukaryotes, the origin recognition complex (ORC) consisting of Orc1-6 is recruited at origins. Eukaryotic origins, contrary to their bacterial and plasmid counterparts, cannot be defined by specific consensus sequences but for the notable exception of *S. cerevisiae*. Nonetheless origins can be defined empirically, for example by using different thymidine analogs to label origins in two successive cell cycles (Patel et al., 2006) or through mapping of ORC1 binding sites using chromatin immuno-precipitations (ChIPs) (Dellino et al., 2013). Genome-wide identification of Okazaki fragments can also pinpoint the location of origins. Indeed, in a landmark work, (Smith and Whitehouse, 2012) exploited a DNA ligase mutant in *S. cerevisiae* to enrich for Okazaki fragments, enabling their identification throughout the genome of the yeast. In vertebrates, rapidly decreasing levels of a protein called geminin (of which there seems to be no orthologues in either budding or fission yeasts) and concomitant increases in levels of Cdt1 enable stepwise recruitment of Cdc6 (Cdc18 in *Sz. pombe*) and Cdt1 at ORC-enriched origins. Importantly, Cdt1 forms a hetero-heptamer with MCM2-7 (minichromosome maintenance) (Tanaka and Diffley, 2002) that is required for loading of MCM2-7 at origins. This enables the sequential loading of two hexamers of MCM2-7 at origins with the double hexamers facing one another head-to-head (Evrin et al., 2009, Gambus et al., 2011, Remus et al., 2009).

The MCM2-7 complex is made of six related but non-identical subunits that are evolutionarily conserved. These genes were identified through the isolation of hypomorphic alleles that were unable to propagate minichromosomes (artificial chromosomes containing a centromere and a single origin of replication) but did not affect the replication of natural chromosomes or minichromosomes with multiple origins (Tye, 1999). The

Mcm proteins are part of the AAA^+ ATPases family (Forsburg, 2004). The MCM2-7 double hexamer loads onto dsDNA but the single hexamer can bind ssDNA with high affinity (Bochman and Schwacha, 2008). Moreover, the MCM2-7 complex has demonstrable helicase activity *in vitro* (although it is a less processive enzyme than the replicative helicase of which it is a constituent) and it has a gate between its Mcm2 and Mcm5 subunits that allows the complex to assume a closed or opened conformation on the loaded DNA (Bochman and Schwacha, 2008).

In eukaryotes, origin licensing and firing are de-coupled to prevent troublesome cases of re-replication and they occur in G_1 and S phase respectively. Entry in S phase is mediated by activation of S phase dependent kinases. In budding yeast, cyclins Clb5 and Clb6 associate with Cdc28 to form the CDKs that promote efficient entry into the mitotic S phase (Epstein and Cross, 1992, Schwob and Nasmyth, 1993). In fission yeast, two S phase cyclins Cig1 and Cig2/Cyc17 associate with Cdc2 to promote the mitotic S phase (Bueno and Russell, 1993, Connolly and Beach, 1994, Obara-Ishihara and Okayama, 1994) although cells can fire origins by using only the mitotic cyclin Cdc13 associated to Cdc2. Defects in S phase progression are observable only upon deletion of the *cig1*, *cig2*/*cyc17* and *cdc13* genes (Fisher and Nurse, 1996).

The CDK phosphorylates several components required for replisome assembly including Sld2 and Sld3-Sld7. This promotes the formation of a complex made up of Sld2 (bound to GINS and Polymerase ε), Sld3-Sld7 (bound to Cdc45) and Dpb11 that is important for productive origin licensing. A second kinase, known as the Dbf4-dependent kinase (DDK), consisting of Dbf4-Cdc7 in *S. cerevisiae* and Dfp1-Hsk1 in *Sz. pombe*, is also critical for replisome assembly. Notably, DDK is responsible for the phosphorylation of MCM2-7 that is a pre-requisite for recruitment of Cdc45 complexed to Sld3-Sld7. Phosphorylated Sld2 is required for the recruitment of GINS. The latter is made up of Sld5, Psf1, Psf2 and Psf3 (GINS: Go, Ichi, Ni, San are the Japanese equivalent of five, one, two and three respectively) (Takayama et al., 2003). This enables the co-localization of Cdc45, GINS and MCM2-7,

leading to the formation of the CMG helicase. Whilst sub-complexes of MCM2-7 and the MCM2-7 hexamer itself have demonstrable helicase activity *in vitro*, the replicative helicase is the CMG (Ilves et al., 2010). Sld2 also recruits Polymerase ε. The CMG associates with Polymerase ε to form the CMGE holoenzyme. The polymerase and helicase within the CMGE co-operate functionally (Langston et al., 2014) but little is known about the role Polymerase ε could play in unwinding of the double helix. The CMG itself is at the centre of a dynamic machinery used to replicate genomes called the replisome. Subunits of the replisome associate with one another with different affinities under physiological conditions.

The exact molecular functions of Cdc45 and of the GINS complex within the CMG are yet to be elucidated. GINS is required for bridging Cdc45 and the MCM2-7 complex and for binding of the CMG helicase to Ctf4 (Gambus et al., 2006, Gambus et al., 2009). *In vitro* analysis indicate that the only subunit of the CMG capable of binding to Ctf4 in isolation is Sld5 (Tanaka et al., 2009a).

Association of GINS and Dpb11 to origins are mutually dependent (Takayama et al., 2003) and the interaction between GINS and Dpb11 is important for efficient initiation at origins (Tanaka et al., 2013). The Dpb2 subunit of Polymerase ε was shown to physically interact with Psf1 in a yeast-two-hybrid assay (Takayama et al., 2003) and *in vitro* assays using the Polymerase ε and GINS purified from *Xenopus* cell extracts revealed that GINS potentiates the catalytic activity of the polymerase (Shikata et al., 2006). Y2H, IPs and *in vitro* reconstitution assays suggest that N-terminus of Dpb2 interacts with the C-terminal B-domain of Psf1 in yeast (Sengupta et al., 2013). This enables tethering of Pol ε to the CMG helicase and, thus, to the rest of the replisome. In fact, when a variant of Dpb2 lacking the N-terminus (first 158 residues) is expressed after depletion of endogenous Dpb2 in G_1-arrested cells, the CMG cannot be assembled and cells are unable to sustain replicative synthesis of DNA (Sengupta et al., 2013). By contrast, depletion of endogenous Dpb2 in G_1-arrested cells followed by expression of Dpb2 (1-168) led to formation of the CMG and progression in S phase once the cells were released from G_1-arrest. Moreover, once the CMG is assembled, Dbp2 but not Dpb2 (159-689)

can associate with the CMG. As such, the N-terminal domain of Dpb2 is required not only for CMG assembly at origins but is also important for Pol ε to load onto on-going forks during S phase, at least in yeast. Both functions require the Psf1 subunit of GINS. Cdc45 is more enigmatic still. It is a homolog of the bacterial RecJ ssDNA exonuclease and is structurally very similar to RecJ in spite of limited similarity in primary sequence (Simon et al., 2016). However, of four active site-motifs, three are inactivated in Cdc45 (Simon et al., 2016). This contributes to the lack of exonuclease activity of the protein *in vitro* (Krastanova et al., 2012). Due to its similarity with RecJ, Cdc45 binds to long ssDNA (80-120 bases long) antagonizing its interaction with the MCM2-7 complex (Bruck and Kaplan, 2013). This may be useful for fork pausing. Cdc45 also interacts with RPA, at least in humans (Szambowska et al., 2017). In gel shift assays, Cdc45 did not bind to ssDNA in isolation but enhanced the binding of RPA to ssDNA. Removal of residues spanning positions 154-164 or 137-188 in Cdc45 led to a complete loss of Cdc45-dependent RPA loading on ssDNA. Cdc45 also binds to branched DNA and mediates limited passive unwinding of 3'-overhanged substrate (Szambowska et al., 2014). The authors of the study go on to suggest that Cdc45 could work as a wedge, facilitating unwinding by the replicative CMG helicase. Cdc45 specifically interacts with the leading strand extended by the CMG helicase. Mutations that abrogate this interaction lead to diminished CMG activity (Petojevic et al., 2015). The authors suggest that Cdc45 prevents slippage of DNA from the opened channel between Mcm2 and Mcm5 when the MCM2-7 is in its opened conformation.

Although the exact contribution of the individual subunits of CMG remains to be determined, the formation of the CMG allows the rearrangement of the inactive MCM2-7 double hexamer (arranged previously head-to-head by virtue of interactions between the N-terminal domains of their constituent proteins) into two 5'-3' DNA helicases. Mcm10 is also required for activation of the CMG helicase. Depletion of Mcm10 does not affect CMG assembly but prevents the CMG from moving along forks (Kanke et al., 2012, van Deursen et al., 2012, Watase et al., 2012). Mcm10 participates in remodelling of the CMG to a functional helicase by either triggering a conformational change in Mcm2,

exposing a buried segment of the protein or by recognizing and binding to this segment (Lõoke et al., 2017). The activated CMG can unwind chromosomal DNA and translocate away from origins. Several models are now proposed for the mechanism of movement of the CMG derived from high resolution cryo-EM images of CMG from yeast and *Drosophila* (Abid Ali et al., 2016).

The unwound DNA is then used as a template for synthesis of new DNA. The enzymes responsible for DNA synthesis are the replicative DNA B-family polymerases: Polymerase α, Polymerase δ and Polymerase ε (see Section 1.3). *In vitro* work in John Diffley's lab indicates that cooperative action between the CMG and DNA polymerases is sufficient for basal replication at slow, non-physiological rates (Yeeles et al., 2017). Moreover, a minimal leading replisome made up of Pol ε, RFC, PCNA, and RPA from *S. cerevisiae* can only achieve rates of 0.26 kb/min on average (Georgescu et al., 2014) while *in vivo* speeds of replisome movement have been measured between 1-2 kb/min (Conti et al., 2007, Hodgson et al., 2007, Sekedat et al., 2010). However, in the presence of the fork protection complex (Csm3, Mrc1 and Tof1) and the clamp loader, PCNA, *in vitro* replisome speeds of up to 1.9 kb/min can be achieved, comparable to physiological speeds (Yeeles et al., 2017). Intriguingly, PCNA would seem to speed up both leading and lagging strand synthesis.

1.3 DIVERSE ROLES FOR POLYMERASE α

DNA Polymerase α, Polymerase δ and Polymerase ε are multi-subunit enzymes required for the duplication of eukaryotic genomes. Their multiple subunits are mostly conserved (Fig 1.2) and the largest subunit from each enzyme is responsible for polymerizing DNA from deoxyribonucleotides (dNTPs). These subunits are all from the B-family polymerases and they adopt a right-hand conformation. In brief, their structure consists of palm domain that houses the catalytic site, the fingers domain that interacts with incoming dNTPs and a thumb domain that holds the nucleic acid substrate close to the catalytic site (Perera et al., 2013).

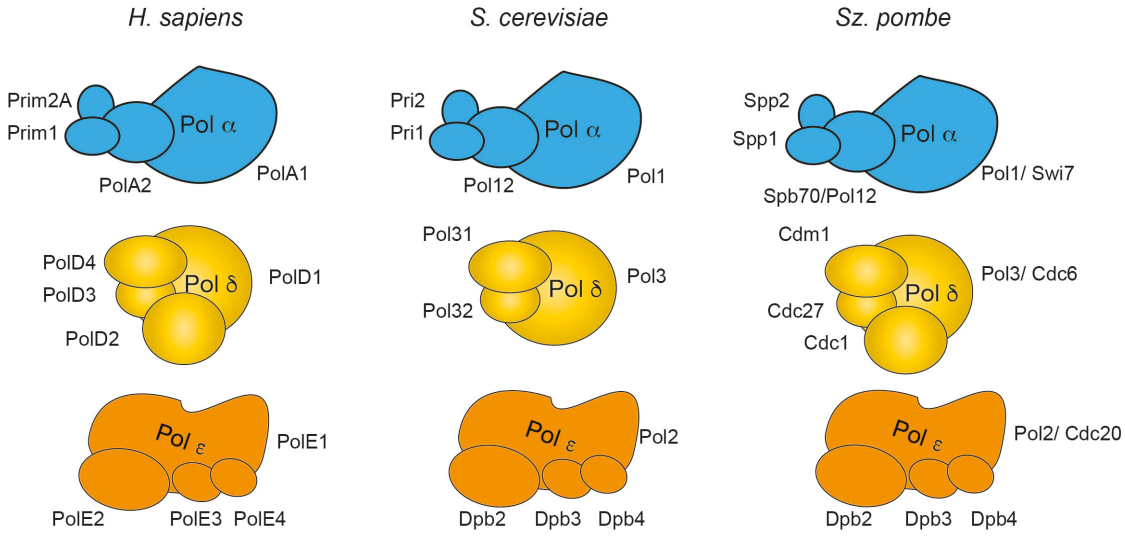


Figure 1.2. The replicative polymerases are multi-subunit enzymes and are conserved across eukaryotes.

Uniquely, Polymerase α also has primase activity mediated by its Pri1 (Spp1 in *Sz. pombe* and Prim1 in humans) and Pri2 (Spp2 in *Sz. pombe* and Prim2 in humans) subunits. The primase is responsible for synthesis of a small stretch of RNA (~10 nucleotides long) called the RNA primer prior to an intramolecular hand-off to the catalytic Pol1 subunit of Pol α (Perera et al., 2013). Synthesis of RNA primers is crucial as DNA polymerases cannot synthesize DNA *de novo*. Indeed, the Pol1 subunit recognizes and binds to the DNA/RNA hybrid left over after priming as the DNA/RNA substrate adopts an A-DNA conformation. Upon synthesis of DNA at the 3' end of the DNA/RNA hybrid, the substrate gradually adopts a conformation between A- and B-DNA, increasingly constraining Pol1's grip as the enzyme translocates to the newly synthesized 3'-end of the substrate until the latter adopts a fully B-DNA conformation and can no longer be used as a substrate by Pol1 (Perera et al., 2013). This feature probably acts as a molecular cue for replication to switch to a more processive enzyme. All the DNA polymerases replicate in a 5' to 3' direction. Consequently, on the leading strand then, Pol α is theoretically needed only once upon origin firing whilst, on the lagging strand, it is continuously required for repeated priming and extension of primers (Fig 1.3). Limited synthesis of DNA by Polymerase α leads to small stretches of DNA that are handed over to a more processive enzyme that go on to synthesize around 200 bases or more. These longer stretches of DNA are known as Okazaki fragments. After extension, these Okazaki fragments are processed by the removal of the RNA primer and their substitution with corresponding DNA segments. Although juxtaposed, the resulting DNA fragments have nicks between their 5' and 3' ends that need to be healed by ligases (Cdc9 and Cdc17 in budding and fission yeast respectively). Deletion of the *POL1*/*pol1* gene is lethal in both budding and fission yeasts (Francesconi et al., 1993) and depletion of Pol1 in fission yeast using the AID degron leads to a tight arrest in early S phase, similar to treatment with HU (Kanke et al., 2011).

In budding yeast, Polymerase α has been shown to interact with Cdc13 and abrogation of this interaction leads to lengthening of telomeres (Sun et al., 2011). Meanwhile, in fission yeast, Pol1 was also found to be necessary for both imprinting at a specific locus in the genome and is involved in silencing

of donor loci (Ahmed et al., 2001, Singh and Klar, 1993, Nakayama et al., 2001a). Both the silencing and imprinting processes are important for cell-type switching in this yeast. This will be discussed further in Section 1.9.

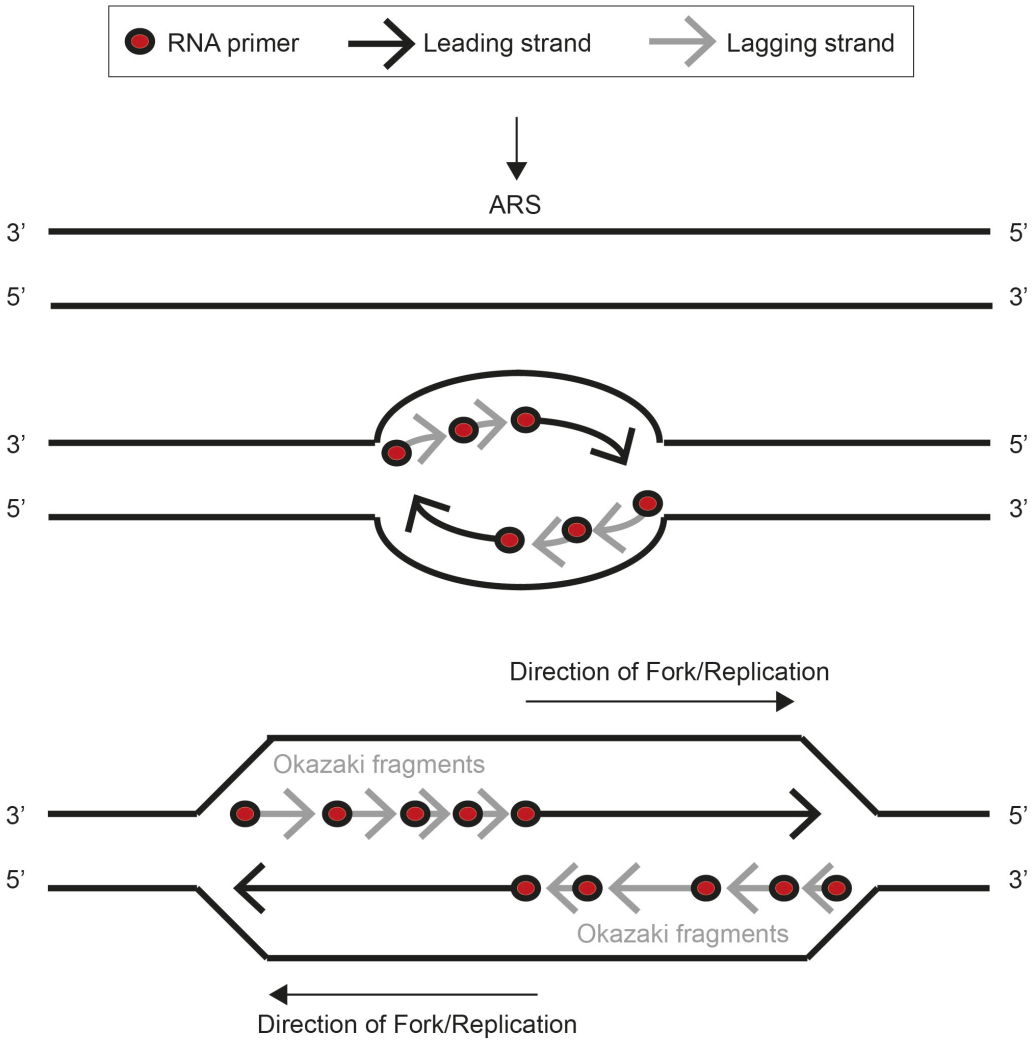


Figure 1.3. Polymerase α is differently enriched on the leading and lagging strands. On leading strands, the enzyme is required at origins. However, because polymerases synthetize DNA in a 5' to 3' direction, lagging strand synthesis occurs discontinuously, so that more molecules of Polymerase α are required on lagging strands.

1.4 THE PROCESSIVE DNA POLYMERASES δ AND ε

After synthesis of the RNA primer and a small stretch of DNA, there is intermolecular hand-off of the nucleic acid substrate from Polymerase α to more processive enzymes, notably DNA Polymerases δ and ε. The catalytic subunits of DNA Polymerases δ and ε are Pol3 and Pol2 respectively. As well as their polymerization site, these proteins also possess 3'-5' exonuclease activities that ensure high fidelity duplication of the DNA template by removing nucleotides incorrectly incorporated into the DNA molecule. Both enzymes then are well suited for processive DNA synthesis but the division of labour between the two of them, based solely on their structures, has been difficult to predict. To determine which of the two proteins synthesized the leading and lagging strands, mutagenic alleles of *POL2* and *POL3* with specific mutation signatures were used to replicate the *URA3* marker gene in *S. cerevisiae*, placed in close proximity to an origin, in both orientations. Using this assay, it was determined that Polymerases ε synthesizes primarily the leading strand whilst Polymerases δ is mainly functional at lagging strands. Results using alleles of the enzymes that encode for variants of the proteins without their exonuclease activities also suggest this particular configuration of labour division. This is reviewed in (Kunkel and Burgers, 2008). A similar experiment was conducted in *Sz. pombe,* using mutagenic *cdc20*/*pol2* and *cdc6*/*pol3* alleles that incorporate elevated levels of ribonucleotides in the genome (Miyabe et al., 2011). Using theses alleles, it was found that Pol ε synthesizes the leading strand of a reporter gene whilst Pol δ synthesizes its lagging strand. This suggests that the division of labour between these two enzymes is conserved evolutionarily (Miyabe et al., 2011). Moreover, a technique was created where the global incorporation of ribonucleotides within the genome of *Sz. pombe* could be assessed (Keszthelyi et al., 2015). Using this method in strains carrying either the mutagenic *cdc20*/*pol2* and *rnh201Δ* or the mutagenic *cdc6*/*pol3* and *rnh201Δ*, it was shown that the Pol ε synthesizes leading strands and Pol δ synthesizes lagging strands throughout the genome of *Sz. pombe* (Daigaku et al., 2015). At efficient origins, however, it was determined that there is a bias for Pol δ to initiate leading-strand replication,

followed by an exchange for Pol ε (Daigaku et al., 2015). Purely *in vitro* work also suggests that Pol δ and ε function at lagging and leading strands respectively (Yurieva and O'Donnell, 2016).

Yet, deletion of the N-terminus of Pol2 in both *S. cerevisiae* and *Sz. pombe* is not lethal (Dua et al., 1999, Feng and D'Urso, 2001). The N-terminus encodes the catalytic activity of the protein, suggesting that Polymerase δ can, at least under such circumstances, replicate both leading and lagging strands in both model organisms. Moreover, in human cells infected with the SV40 virus, Pol δ in conjunction with Pol α, but not Pol ε were found to crosslink to the viral DNA (Waga et al., 1994). Some genetic evidence also seems to indicate that Pol δ is the major replicative polymerase that synthesizes both leading and lagging strands of DNA whilst Pol ε is mostly involved in correcting mistakes on the leading strand (Johnson et al., 2015). Nevertheless, the current consensus is that Pol ε and Pol δ synthesize the leading and lagging strands respectively.

Both DNA Polymerases δ and ε are known to bind to PCNA (Chilkova et al., 2007). PCNA itself is a conserved protein that forms a homotrimeric complex that acts as a clamp by encircling DNA (Krishna et al., 1994, Yao et al., 1996). It is loaded onto DNA by RFC1 (replication factor C) and the RFC2-5 complex (Majka and Burgers, 2004). PCNA also binds to a number of proteins and, in doing so, it tethers these proteins to the DNA. There are differences between the interaction of these two polymerases with PCNA. Surface plasmon resonance studies have revealed that Pol ε interacts weakly with PCNA whilst Pol δ interacts strongly with it. This is offset by the fact that Pol ε has a strong affinity for dsDNA, ssDNA and primed DNA whilst Pol δ does not. *In vitro* primer extension assays also demonstrated that binding to PCNA stimulates Pol ε processitivity ~6-fold whilst binding to PCNA stimulates Pol δ processitivity at least ~100 fold (Chilkova et al., 2007), suggesting that PCNA is critical for the functioning of Pol δ.

Proteins that interact with PCNA are known to do so through their respective PIP box (PCNA interacting peptide) (Warbrick, 1998). Besides polymerases, other proteins also encode PIP boxes, including the Flap endonucleases Fen1 and the DNA damage protein Rad2 (Yu et al., 2014). In yeast, the Pol2 subunit of Pol ε contains a PIP box but point mutagenesis of the motif does not affect growth, suggesting that the putative physical interaction between Pol2 and PCNA is not essential for Pol ε's functions during replication (Dua et al., 2002). However, any difference in binding-affinity to PCNA between wildtype Pol2 and its PIP box mutants was not tested (Dua et al., 2002). Interestingly, unlike both Pol α and δ, Pol ε has a P domain that allows it to encircle dsDNA on its own (Hogg et al., 2014). Evolution of this novel P domain might perhaps compensate for the inactivation of the PIP box within Pol ε. PCNA, however, is still important to achieve maximal rates during leading-strand synthesis (Yeeles et al., 2017). Meanwhile, in Pol δ, the Pol32 subunit contains a PIP box whilst a cysteine motif (CysA) in the Pol3 subunit also enables direct interaction with PCNA (Johansson and Dixon 2013, Netz et al., 2011). This dual method of Pol δ to interact with PCNA highlights the functional importance of this interaction during DNA replication.

The inner surface of the PCNA clamp is lined with lysine- and arginine-rich α-helices and it is ~35 Å wide whilst the double helix which it encircles is ~24 Å wide (Wing et al., 1980). The DNA is threaded through PCNA at a 15° tilt (De March et al., 2017). Crystallographic data suggest that five basic residues (K20, K77, R149, H153 and K217) from one of the PCNA isomer as well an additional basic residue (K80) on the adjacent isomer within the same trimer form polar interactions with five consecutive phosphate in one strand of the DNA backbone (De March et al., 2017). Simulations also suggest that the PCNA trimer can stably interact with both strands of DNA where the protein will rotate in a cogwheel-like fashion shifting interactions between its residues from antecedent or succeeding phosphates on the DNA backbone to the phosphates immediately adjacent (De March et al., 2017). As such, when in this coupled state to DNA, the PCNA trimer will move by 1 base pair at a time, contrary to its uncoupled state where it freely slides over DNA (Yao et al.,

1996). The coupled state enables Pol δ to translocate quickly enough for processive synthesis of DNA.

1.5 OTHER PROTEINS THAT ASSIST DNA REPLICATION

The CMG is at the centre of a dynamic assembly of proteins known as the replisome. The identity of the proteins within the replisome were initially identified by consecutive immuno-precipitation of GINS and MCM2-7 followed by mass spectrometric analysis of the immuno-precipitated proteins (Gambus et al., 2006, Gambus et al., 2009). Given the nature of screens however, it is likely that other replisome components exist. Table 1.1 lists some of the replisome proteins identified. For brevity, I will focus on components of the fork protection complex (Csm3, Mrc1 and Tof1), as well as Ctf4 that forms a homotrimeric complex and works as a hub to tether proteins to the CMG helicase (Simon et al., 2014). As such, the recruitment of accessory replisome components can be thought as either Ctf4-dependent or Ctf4-independent.

Table 1.1. Accessory proteins of the replisome

Identity	Orthologue (*Sz. pombe*)	Cft4-dependent
Ctf4	Mcl1	Not applicable.
Top1	Top1	No
Tof1	Swi1	No. Physically interacts with Top1. Also interacts with phosphorylated CMG and MCM2-7. GINS dependent.
Csm3	Swi3	No
Mrc1	Mrc1	No
Mcm10	Mcm10	No (Via Pol α and MCM2-7)
Spt16	Spt16	No
Pob3	Pob3	No
Dna2	Dna2	Yes
Tof2	Not identified	Partially. Also interacts with Top1.
Dpb2	Dpb2	Partially. (Note: Dpb2 is a subunit of Pol ε but it can also function as an accessory protein). Also interacts with Psf1.
Chl1	Chl1 (predicted)	Yes.
Dia2	Pof3	Partially. Also interacts with Mrc1.

Mms22	Mus7	Yes. Interacts with WD40 domain.

Ctf4 was originally isolated as a binding partner of Pol1 in *S. cerevisiae* (Miles and Formosa, 1992b). Both Ctf4 and its interaction with Pol1 are conserved as seen with the *Sz. pombe* orthologue, Mcl1 (Williams and McIntosh, 2002, Williams and McIntosh, 2005) as well as with the vertebrate orthologue, AND-1 (Zhu et al., 2007). Interestingly, whilst *ctf4*Δ is viable in *S. cerevisiae*, *mcl1*Δ is conditionally lethal in *Sz. pombe* and the depletion of the vertebrate orthologue of Ctf4 (AND-1) leads to incomplete DNA replication in *Xenopus* (Zhu et al., 2007) and in humans (Bermudez et al., 2010). Besides a role in DNA duplication and metabolism, Ctf4 is also required for genome stability and for sister chromatid cohesion (Bermudez et al., 2010, Gosnell and Christensen, 2011, Hanna et al., 2001, Kouprina et al., 1992, Miles and Formosa, 1992a, Tanaka et al., 2009b, Williams and McIntosh, 2002, Zhou and Wang, 2004, Zhu et al., 2007).

In yeast, Ctf4 is made up of three subunits, consisting of a WD40 domain in its N-terminus spanning residues 2-383 (Gambus et al., 2009), a six-bladed β-propeller domain as well as an α-helical bundle at its C-terminus (Simon et al., 2014). Noteworthy is a second WD40 predicted in the C-terminal half of the protein (Simon et al., 2014). Crystallographic data have revealed that Ctf4 trimerizes by virtue of its central β-propeller domain. This trimer is soluble and both the WD40 and α-helical domains extend away from the plane formed by the β-propeller trimer, primed for interaction with a host of proteins and perhaps with DNA (Simon et al., 2014). Ctf4 interacts with both Pol1 and Sld5 by virtue of its C-terminus (Gambus et al., 2009, Tanaka et al., 2009a). The corresponding domain in AND-1 is also required for interaction with PolA1 but not does not interact with human Sld5 *in vitro* (Guan et al., 2017). Both Pol1 and Sld5 requires a CIP box (Ctf4 Interacting Peptide) to interact with Ctf4 (Simon et al., 2014). The CIP box is composed of the 'DDIL' residues at its core. Two other proteins have been shown to interact with Ctf4 through their CIP box; the flap endonuclease Dna2 that is important in the processing of Okazaki fragments (Villa et al., 2016) and the Chl1 DNA helicase that is

important for sister chromatid cohesion (Samora et al., 2016). A second group of proteins with a divergent CIP box also bind to Ctf4 (Simon et al., 2014). These include the Dpb2 subunit of Polymerase ε and the *rDNA*-associated protein Tof2. Crystallographic data has shown that the two types of CIP boxes bind to different residues within the C-terminus of Ctf4. Whether all physical interactors of Ctf4 require one of the two types of CIP box however is not known but improbable. By virtue of forming a trimer and through its interaction with Sld5, Ctf4 acts as a hub that tethers several proteins to the CMG helicase, including Pol1. Deletion of *CTF4* prevents Pol1 from binding to other components of the replisome in S phase and mutation of CIP box of Pol1 also abrogates interaction between Pol1 and the replisome (Gambus et al., 2009, Simon et al., 2014). This suggests alternative mechanisms that allow Pol α to coordinate with forks. Alternatively, Pol α simply needs not coordinate with forks.

The fork protection complex (FPC) is a sub-complex of the replisome (Bando et al., 2009). It is made up of Csm3, Mrc1 and Tof1. Tof1 was originally identified as an interactor of the topoisomerase/gyrase Top1 (Park and Sternglanz, 1999). All three proteins are conserved. In *Sz. pombe*, the orthologues are known as Mrc1, Swi1^{Tof1} and Swi3^{Csm3} whilst in humans they are known as Claspin^{Mrc1}, Tim^{Tof1} and Tipin^{Csm3}. The association of Csm3 and Tof1 to the replisome is co-dependent whilst Mrc1 requires both Csm3 and Tof1 to associate efficiently with replisomes (Bando et al., 2009). Conversely, Mrc1 is not required for either Csm3 or Tof1 to bind to other components of the replisome. In *Sz. pombe*, Swi1^{Tof1} and Swi3^{Csm3} form a complex (Noguchi et al., 2004) and are mutually dependent for stability (Shimmoto et al., 2009). Mrc1 interacts with Swi1 and Swi3 independently of Swi3 and Swi1, respectively. The difference between the Mrc1 orthologues of these two yeasts can be attributed to a conserved helix-loop-helix motif in the DNA-binding domain of *Sz. pombe* Mrc1 that is not present in the *S. cerevisiae* orthologue (Zech et al., 2015) so that this motif of Mrc1 is exploited for formation of the FPC in fission yeast.

The FPC is required for efficient replication barrier activities in several organisms, including in both budding and fission yeasts as well as in humans. For instance, it is required for pausing and stalling at the *MPS1* and *RTS1* loci in *Sz. pombe*. These are pre-requisite for cell-type switching in the organism. The FPC is also required for fork arrest prior to entry at the highly-transcribed *rDNA* loci. This is a conserved function, first identified in budding yeast (Mohanty and Bastia, 2004, Kobayashi, 2003) but later expanded to other organisms. This is reviewed in (Bastia and Zaman, 2014). The FPC is also required for efficient arrest at fork barriers at centromeres and telomeres and is involved in sister chromatid cohesion (Xu et al., 2007, Smith-Roe et al., 2011).

Interestingly, Mrc1 is also involved in the intra-S phase checkpoint. In the event of replication stress such as fork stalling as a result of obstacles or depressed pools of dNTPs (as seen upon hydroxyurea-treatment), cells activate a signal transduction pathway that delays exit from S phase until cells completely replicate their DNA (Allen et al., 1994, Weinert et al., 1994). This is known as the intra-S phase checkpoint. Besides delaying entry into mitosis, this checkpoint delays firing of late origins by hyper-phosphorylating the Dbf4 component of DDK and by phosphorylating Sld3 (Lopez-Mosqueda et al., 2010, Weinreich and Stillman, 1999, Zegerman and Diffley, 2010) and causes global changes in transcriptional activity to promote DNA repair and to reduce replication stress (Gasch et al., 2001), such as in the *CRT1* gene, an inhibitor of transcription of the *RNR* genes in *S. cerevisiae* that are responsible for synthesis of dNTPs from rNTPs (Huang et al., 1998). The intra-S phase checkpoint also serves the purpose of preventing damage-induced fork catastrophe (Desany et al., 1998, Lopes et al., 2001, Tercero and Diffley, 2001). The intra-S phase checkpoint involves an upstream checkpoint kinase, Mec1 (Rad3 in *Sz. pombe* and ATR in humans). Meanwhile, the effector downstream kinase Rad53 (Cds1 in *Sz. pombe* and Chk1 in humans) is activated by hyperphosphorylation in response to DNA damage and stress (Iyer and Rhind, 2017).

Mrc1 is a known mediator of the intra-S checkpoint and works at an intermediate step between Mec1 and Rad53 but is phosphorylated by both kinases (Alcasabas et al., 2001, Osborn and Elledge, 2003, Tanaka and Russell, 2001). Mutations in the *MRC1* allele can cause delayed but persistent Rad53 activation and further deletion of the *RAD9* gene (a mediator that activates Rad53 in response to DNA damage) completely abrogates Rad53 phosphorylation. Depending on the background used, deletion of both *MRC1* and *RAD9* is lethal but this can be suppressed by overexpression of *RNR1*, similar to *rad53*Δ (Alcasabas et al., 2001, Redon et al., 2006). As such, Mrc1 is required for Rad53 phosphorylation in response to DNA stress and, when this pathway is impeded, Rad53 is phosphorylated solely in response to DNA damage occurring behind replication forks. This signalling pathway is conserved in fission yeast (Zhao et al., 2003). Triple *cdc13-1 mrc1*Δ *tof1*Δ mutants are characterized by pronounced growth defects, not associated with either the *cdc13-1 mrc1*Δ *or cdc13-1 tof1*Δ double mutants, highlighting an additive effect of the *MRC1* and *TOF1* deletions as checkpoint mediators (Grandin and Charbonneau, 2007). Similarly, in *Sz. pombe*, both *swi1*Δ and *swi3*Δ (Noguchi et al., 2003, Noguchi et al., 2004) lead to a pronounced but comparable decrease in Cds1 activation upon HU addition whilst *mrc1*Δ reduces Cds1 activation beneath detectable levels in HU-treated cells (Tanaka and Russell, 2001).

Taken together, the three components of the FPC (Csm3, Mrc1 and Tof1) have distinct roles, especially as mediators of cell cycle checkpoints where Mrc1 has a more prominent role. However, they do share some overlapping functions such as in stabilizing stalled forks upon encounters with replication barriers. These barriers occur throughout the genome and can either be pre-programmed or arise stochastic. Fork barriers can also be made up completely of nucleic acids or be composed of nucleic acid-protein complexes. The following chapters will be devoted to these fork barriers, how they challenge the process of DNA replication, how forks respond to them and how cells can actually use such barriers to their advantage.

1.6 BARRIERS TO THE REPLICATION PROCESS

During DNA replication, forks do not proceed freely through genomes but instead encounter several barriers that can cause fork stalling or even replication termination. Some replication fork barriers (RFBs) consist of nucleic acids. Often, they will form secondary structures that physically impede fork progression. A subset of such nucleic acid fork barriers is made of both DNA and RNA and arises during transcription. These are known as R-loops. Other fork barriers are composed of DNA-protein complexes. Interestingly, cells seem to exploit such barriers at pre-programmed locations to mediate useful cellular events such as pausing at the *rDNA* loci.

1.6.1 NUCLEIC ACID BARRIERS

There are several subtypes of nucleic acid barriers (Fig 1.4) (Mirkin and Mirkin, 2007). One of these are hairpins that arise because of inverted repeats. Seminal work from the DePamphilis lab has shown that phage DNA encoding sequences that lead to hairpin formation can promiscuously cause replication termination of Pol α from CV-1 cells (derived from monkey kidney tissue), HeLa cells and calf thymus tissue *in vitro* (Weaver and DePamphilis, 1982). These sequences also terminated replication driven by the T4 DNA polymerase. In fact, the authors estimated that 28% of replication arrests were attributed to such hairpin structures. Similarly, in a study using Pol α extracted from *D. melanogaster* embryos to replicate murine mtDNA cloned into the single-stranded M13 vector, it was shown that hairpins structures can cause replication stalling (Kaguni and Clayton, 1982).

G-quadruplexes represent a different subtype of nucleic acid barriers. They consist of four guanine bases where each guanine forms hydrogen bonds with two other guanine bases. The structure is stabilized by cations, chiefly K^+. G-quadruplexes have also been shown to hinder the replication progression. Indeed, *in vitro* assays where G-quadruplexes are stabilized (in the presence of K^+ ions) cause replication stalling but this pausing is alleviated when the K^+

ions is substituted with Na^+ or NH_4^+ or when cations are removed completely (Usdin and Woodford, 1995, Woodford et al., 1994, Weitzmann et al., 1997).

A third type of nucleic acid fork barrier is the triplex H-DNA that arises because of mirror repeats within DNA. (Mirkin and Mirkin, 2007). The barrier activity of H-DNA was demonstrated using *in vitro* assays where $(TCTCTC)_n$ and $(GAGAGA)_n$ mirror repeats tracts were cloned into the M13 vector. Replication of these substrates by either calf thymus Pol α, the Klenow fragment or Taq polymerase was prematurely terminated in the middle of the individual tracts. However, either using analogs of dNTPs unable to form hydrogen bonds with other bases or pre-incubating templates with single-stranded DNA binding protein from *E. coli* prevented the premature replication termination of the template (Baran et al., 1991, Lapidot et al., 1989). This is consistent with a mechanism whereby the single-stranded part of the template folds back and anneals to the newly synthesized DNA, sand-witching and effectively trapping the polymerase extending the mirror repeat.

Yet another type of nucleic species that hinders fork progression are the DNA/RNA hybrids. These include R-loops that will be the focus of future sections in this chapter. DNA/RNA hybrids also include RNA primers synthesized by the primase subunit of Pol α hybridized to their template (Joyce, 1997, Nick McElhinny et al., 2010a, Nick McElhinny et al., 2010b). So far, no reports have yet implicated these RNA primers as barriers to replication forks, perhaps on account that they are efficiently removed by the flap endonucleases Fen1 and Dna2 (Rossi and Bambara, 2006). However, primer release on the lagging strand is rate-limiting in the T7 replisome-mediated lagging strand synthesis (Hernandez et al., 2016) whilst T7 primase has been shown to act as a brake to prevent leading-strand synthesis from synthesizing exceedingly ahead of the lagging strand (Lee et al., 2006). In neither study however have RNA primers themselves been shown to physically impede fork progression and it is unlikely that RNA primers can obstruct fork progression as *in vitro* DNA synthesis by Pol δ on naked DNA is essentially unrestricted and the polymerase efficiently displaces the RNA primer in the process (Devbhandari et al., 2017).

Ribonucleotides incorporated into the genome constitute yet another type of DNA/RNA hybrid that can act as fork barriers. Indeed, whilst replicative polymerases can polymerize past a ribonucleotide incorporated within the template DNA *in vitro*, a subpopulation of forks will terminate at those ribonucleotides (Watt et al., 2011). As for ribonucleotides inserted in the genome *in vivo*, they are efficiently removed by the RNase H2 endonuclease (Nick McElhinny et al., 2010b). Absence of RNase H2 leads to genome-wide instability in yeast (O'Connell et al., 2015). Moreover, yeast strains carrying both the *pol2-M644G* (an allele that encodes for a Pol2 variant that incorporates ribonucleotides in the genome 11-times more frequently than its wildtype counterpart) and the *rnh201*Δ alleles (a deletion mutant of one subunit of RNase H2) progress slowly in S phase. The triple *pol2-M644G rnh1*Δ *rnh201*Δ mutant is not viable, suggesting that RNase H1 is also involved in removal of individual ribonucleotides inserted into the genome (Lazzaro et al., 2012). The *rnh1*Δ *rnh201*Δ mutants are characterized by hyper-phosphorylated Rad53 upon hydroxyurea treatment, indicative of replication stress (Lazzaro et al., 2012). As such, ribonucleotides can constitute a barrier to replication forks. In fact, in *Sz. pombe*, two ribonucleotides incorporated at a particular locus in the genome (*MPS1*, mat1-pausing site 1) is thought to stall replication, leading to fork reversal and, eventually, to synthesis-dependent strand annealing (Vengrova and Dalgaard, 2004, Vengrova and Dalgaard, 2006). This contributes to cell-type switching in this yeast.

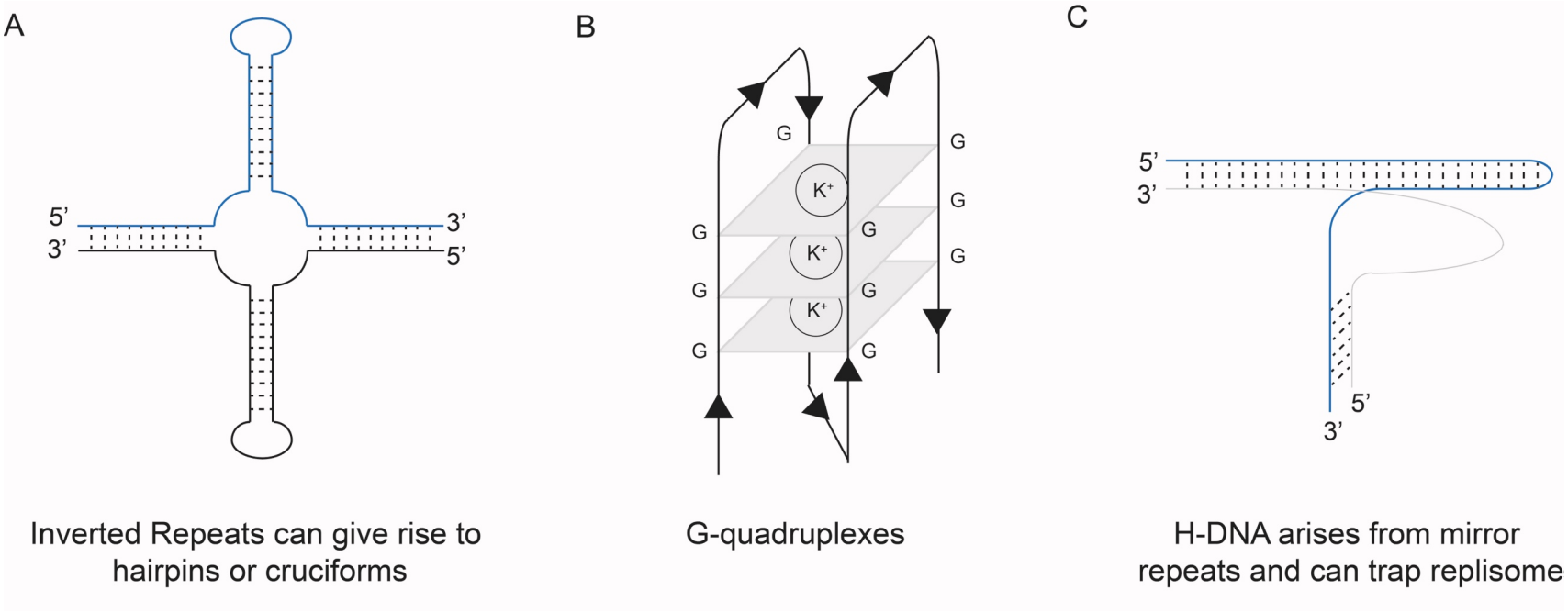


Figure 1.4. Nucleic acid fork barriers assume secondary structures that are refractory to fork progression. Several non-B DNA nucleic acid can act as fork barriers. These include, but are not limited to: **(A)** Inverse repeats within the genome that can form hairpins and cruciform structures, **(B)** G-rich sequences that can be appropriate substrates for the formation of G-quadruplexes, **(C)** Mirror repeats that give rise to H-DNA that can trap replisomes during DNA synthesis.

1.6.2 DNA-PROTEIN COMPLEXES AS REPLICATION FORK BARRIERS

Several DNA-protein complexes that act as barriers to forks have also been identified and these complexes have diverse roles. In prokaryotes, replication termination depends on such DNA-protein complexes. Replication begins at a single origin (*oriC*) and proceeds in a bi-directional manner, halving the genome in two replichores. In theory, because of the different contents of each replichore in terms of DNA accessibility and the stochastic nature of DNA damage, the two leading strands (and their associated replisomes) need not meet at the opposite pole to the *oriC* and the forks can fuse at any place other than the *oriC* itself. In actuality, this scenario is averted by the presence of replication termination sites (Ter) that form a replication trap where forks can travel past a point in one direction (permissive end) but not in the opposite direction (non-permissive end). In *E. coli*, monomeric *trans*-acting Tus proteins (terminus utilization substance) bind to each of ten 23 bp Ter sites (TerA-J) arranged in two opposite orientations and these Ter-Tus DNA-protein complexes somehow enable polar arrest of replication by counteracting the activity of the replicative helicase (DnaB) in an orientation-specific manner (Fig 1.5A) (Mirkin and Mirkin, 2007). Three different mechanisms have been suggested to account for this polar termination. Firstly, the crystal structure of the Tus protein complexed to Ter DNA was obtained and revealed that the Tus protein has two protruding α-helical regions and forks stall when approaching on the side of those protrusions (Kamada et al., 1996). The Tus protein has another asymmetric feature, notably its L1 loop, and forks stall when approaching on the side of this loop. Thus, this L1 loop could also contribute to polar fork termination.

An alternative or complementary model originated from the observations that, although binding to the Ter sites was absolutely necessary for replication arrest, there were some point mutants of TerB to which the Tus protein retained wildtype affinity but where those TerB mutants were nonetheless characterized by defects in replication termination *in vivo* (Coskun-Ari and Hill, 1997). Moreover, the interaction between Tus and TerB was differently affected by mutations depending on their location relative to the site of fork

arrest (Neylon et al., 2000). In fact, *in vitro* experiments have shown that the helicase activity of the DnaB helicase causes efficient dissociation between Tus and TerB when approaching from the permissive end of the arrest site. But, when the DnaB helicase approaches from the non-permissive end, it causes a newly unwound cytosine to flip and bind within a cleft within the Tus protein (Mulcair et al., 2006). Importantly, this cytosine is stringently conserved in the different Ter sequences. In the flipped base configuration, the DnaB would seem to be unable to displace the Tus protein anchored to its Ter site on the DNA, leading to polar fork arrest. Interestingly, cloning arrays of TerB as well as the Tus gene (under the *GAL1* promoter) in *S. cerevisiae* leads to a polar pausing event in S phase that is independent of either Tof1 or Rrm3. Contrary to *E. coli*'s DnaB however, the CMG helicase is not permanently arrested at the ectopically-cloned Ter sites (Larsen et al., 2014).

In *B. subtillis*, a similar mechanism is used to prevent collisions between two replisomes. The *B. subtilis* genome carries nine 29 bp Ter sites (TerI-IX) located at the opposite pole to its *oriC*. These Ter sequences do not share sequence homology with *E. coli*'s Ter sites nor are they predicted to adopt similar secondary structures (Wilce et al., 2001). The Ter sequences are conserved amongst themselves and consist of two functional sites (A and B) to which their cognate binding partner can bind. This protein is known as the replication terminator protein (RTP). RTP shares no primary, secondary or tertiary similarity with *E. coli*'s Tus protein. In *B. subtilis*, two copies of RTP form a dimer that binds to both functional sites within the Ter sequence. Should incoming forks travel proximal to the A site, the fork will be allowed to proceed unhindered whilst forks travelling proximal to the B site will be arrested (Fig 1.5B) (Griffiths et al., 1998, Vivian et al., 2007). It would seem that these two similar systems have evolved separately in *E. coli* and *B. subtilis* (Wake, 1997). As such, it can be theorized that restricting replication termination to a small region of the chromosome is beneficial in organisms with circular genomes. Indeed, in *E. coli*, the *dif* site is located within this region of the genome. This locus is required for decatenation and efficient segregation of sister chromatids, enabling a smooth transition between DNA replication and cell division (Wake, 1997). It is also possible that restricting the

site of replisome collision/fusion to a relatively small, gene-sparse region of the genome is beneficial for chromosomal stability.

1.6.3 DNA-PROTEIN REPLICATION BARRIERS IN EUKARYOTES

In eukaryotes, the best studied example of RFBs are the ones found within non-transcribed spacer regions between *rDNA* repeats, regions of DNA that encode for the 25-28S, 16-18S and the 5.8S rRNA that constitute ribosomes. These were initially discovered in *S. cerevisiae* where replication intermediates of the ~200 tandem repeats of rDNA on chromosome XII were visualized on 2D-gels (Brewer and Fangman, 1988). The 2D-gels demonstrated that forks approaching distally from the promoter from which RNA Pol I transcribes the *rDNA* genes were arrested in a polar fashion. However, forks approaching proximally were not, so that replication of the *rDNA* locus only occurs in one orientation (Fig 1.5C). The barrier activity was found to depend on a *cis*-acting sequence within a non-transcribed region known as *RFB* and on a *trans*-acting factor, Fob1 (Kobayashi and Horiuchi, 1996). Within the *RFB* locus, there exist three discrete sites *RFB1*, *RFB2* and *RFB3*. *RFB1* is sufficient for fork arrest whilst *RFB2* and *RFB3* co-operate as a second, minor fork arrest site (Kobayashi, 2003). It has been suggested that the roles of these barriers is to prevent head-on collisions between the replication machinery and transcribing complexes. Indeed, in a *FOB1*-defective strain where the copy number of rDNA genes was also reduced to around 20, increased collisions between the replication fork and transcribing complexes eventually led to production of extrachromosomal *rDNA* circles and shortened lifespans (Takeuchi et al., 2003). Two other factors are essential for fork arrest at the Fob1-*RFB* site. These are Tof1 and Csm3. Deletion mutation of either gene leads to loss of pausing signal on 2D gels at the *RFB* (Calzada et al., 2005). On the other hand, the third component of the FPC, Mrc1, is dispensable for fork arrest at *RFB*.

Replication barriers at rDNA genes have also been discovered in other species, including in peas (López-Estraño et al., 1999), the ciliate *Tetrahymena thermophile* (MacAlpine et al., 1997), *Xenopus* (Maric et al.,

1999) and humans (Little et al., 1993). Interestingly, the human RFB is non-polar and forks are paused in both directions, as seen on 2D gels (Little et al., 1993). Replication arrest occurs at multiple (up to five) Sal box elements (T1-T5) to which the protein TTF-1 binds (transcription termination factor 1) to mediate arrest. The T2 and T3 boxes are inactivated and it was found that TTF-1 does not bind to either of them nor can they mediate fork arrest (Pfleiderer et al., 1990). The T1 box mediates fork arrest distally from the *rDNA* promoter. Whilst individual Sal boxes do not demonstrate fork arrest proximal to the promoter, T4 and T5 cooperatively caused fork arrest in either direction, suggesting that they act with one another to form a heterochromatic region that impedes fork progression, leading to the peculiarity of non-polar fork arrest (Akamatsu and Kobayashi, 2015). Knockdown of Tim^{Tof1} by siRNA correlated with reduced barrier activity, suggesting a conserved function of the FPC in efficient fork arrest (Akamatsu and Kobayashi, 2015). Interestingly, TTF-1 also participates in transcription termination.

In *Sz. pombe*, there are four fork barriers at the *rDNA* loci (*RFB1-4*). *RFB4* seems to play a minor role unless the barrier activity of *RFB1-3* are compromised (Krings and Bastia, 2004). The *Sz. pombe* orthologue of TTF-1, Reb1, has been shown to bind to *RFB2* and *RFB3* and *reb1*Δ abolishes fork arrest at both sites (Sánchez-Gorostiaga et al., 2004). *RFB1* is the strongest arrest site, to which a homodimer of the essential Sap1 protein binds (Krings and Bastia, 2005). *RFB1-3* also requires the activity of all three components of the FPC (Swi1, Swi3 and Mrc1) for optimal barrier activity (Krings and Bastia, 2004, Zech et al., 2015). As mentioned earlier, Mrc1 from *Sz. pombe* contains a conserved helix-loop-helix motif that may help it to contribute to the fork barrier activity. Interestingly, the FPC is also required for optimal barrier at two other loci, namely at the replication termination site 1 (*RTS1*) and at the *mat1*- pausing site 1 (*MPS1*). Notably, at *RTS1*, two other *trans*-acting factors, Rtf1 and Rtf2 are also required (Codlin and Dalgaard, 2003). Rtf1 is a paralog of Reb1 as well as an orthologue of mammalian TTF-1 (Eydmann et al., 2008). Fork arrest at *RTS1* and fork pausing at *MPS1* are critical for changing the gene expressed at the mating-type locus, *mat1*. This leads to a parent of a particular cell-type (or mating-type) having progenies of different cell-types.

This pattern of mitotic division is known as **asymmetric cell division**. This will be described in more detail in Section 1.7.

Several other replication fork barriers (RFBs) have been described and have been reviewed elsewhere (Dalgaard et al., 2011, Gadaleta and Noguchi, 2017, Labib and Hodgson, 2007, Mirkin and Mirkin, 2007). These include RFBs present at centromeric and telomeric loci. Centromeric RFBs lead to non-polar replication stalling in early S phase again seemingly to prevent conflicts between DNA replication and transcription. In addition, telomeric DNA at the ends of chromosomes are difficult to replicate by virtue of their repetitive nature as well as other local features such as the propensity to form T-loop structures. In *Sz. pombe*, the repetitive nature of the telomeric DNA leads to fork stalling. Pausing may be required for the stabilization of telomeres as abolition of pausing in *swi1*Δ strains leads to loss of the telomeric repeats (Gadaleta et al., 2016).

One interesting feature of pre-programmed fork barriers is that they seldom lead to profound genome instability and that cells do not activate any checkpoint in response to forks stalling at such barriers. Instead, cells seem to respond with limited and controlled instability before resolution of the conflict (Dalgaard et al., 2011). For instance, in *S. cerevisiae*, neither the absence of *MEC1* nor that of *RAD53* affects the stability of replisomes at stalled forks (Calzada et al., 2005). Similar conclusions can be derived from experiments conducted in *Sz. pombe*. In one study, two copies of fork arrest site (*RTS1*) were cloned on either side of a *ura4* reporter gene (Lambert et al., 2005). Importantly, resolution of fork arrests at this construct required proteins involved in homologous recombination but not proteins involved in checkpoint function (including $Cds1^{Rad53}$, Chk1 or $Rad3^{Mec1}$). That is not to say that pre-programmed RFBs cannot be dangerous to cells. For instance, the TerB in *E. coli* when cloned in otherwise stable plasmid vectors were found to be a site of deletion hotspot, leading to progeny plasmids smaller than their parental plasmid (Bierne et al., 1991). Thus, there must be a mechanism that counteract this deletion hotspot characteristic of TerB in the *E. coli* genome to prevent loss of genetic material.

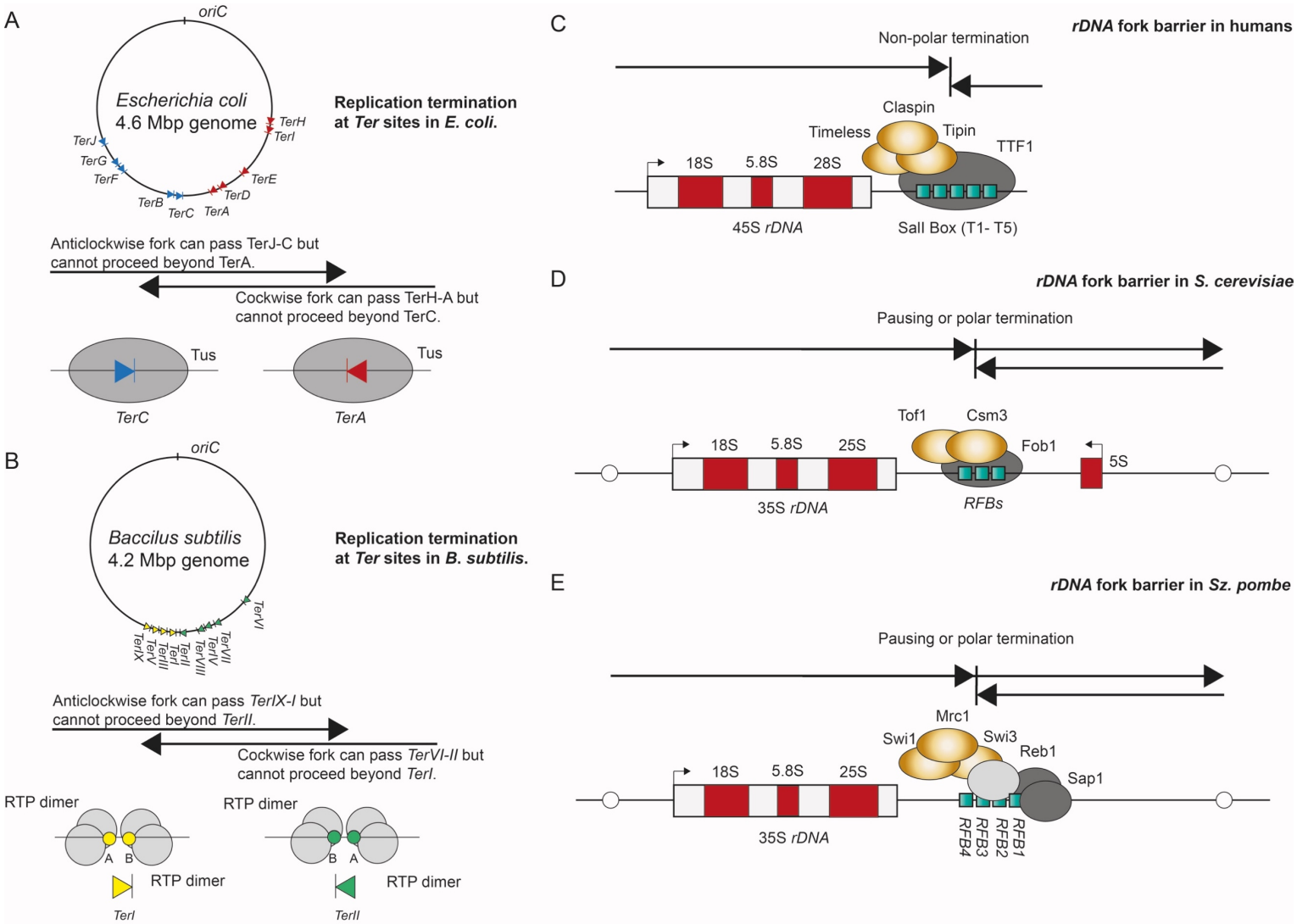

A
oriC
Escherichia coli
4.6 Mbp genome
TerH
TerI
TerJ
TerG
TerF
TerE
TerD
TerB
TerC
TerA
Replication termination at Ter sites in E. coli.
Anticlockwise fork can pass TerJ-C but cannot proceed beyond TerA.
Cockwise fork can pass TerH-A but cannot proceed beyond TerC.
Tus
Tus
TerC
TerA
B
oriC
Baccilus subtilis
4.2 Mbp genome
TerVI
TerX
TerV
TerIII
TerI
TerII
TerVIII
TerIV
TerVII
Replication termination at Ter sites in B. subtilis.
Anticlockwise fork can pass TerIX-I but cannot proceed beyond TerII.
Cockwise fork can pass TerVI-II but cannot proceed beyond TerI.
RTP dimer
RTP dimer
A B
B A
RTP dimer
RTP dimer
TerI
TerII
C
rDNA fork barrier in humans
Non-polar termination
Claspin
Timeless
Tipin
TTF1
18S
5.8S
28S
45S rDNA
Sall Box (T1- T5)
D
rDNA fork barrier in S. cerevisiae
Pausing or polar termination
Tof1
Csm3
18S
5.8S
25S
Fob1
5S
35S rDNA
RFBs
E
rDNA fork barrier in Sz. pombe
Pausing or polar termination
Mrc1
Swi1
Swi3
Reb1
Sap1
18S
5.8S
25S
35S rDNA
RFB4
RFB3
RFB2
RFB1

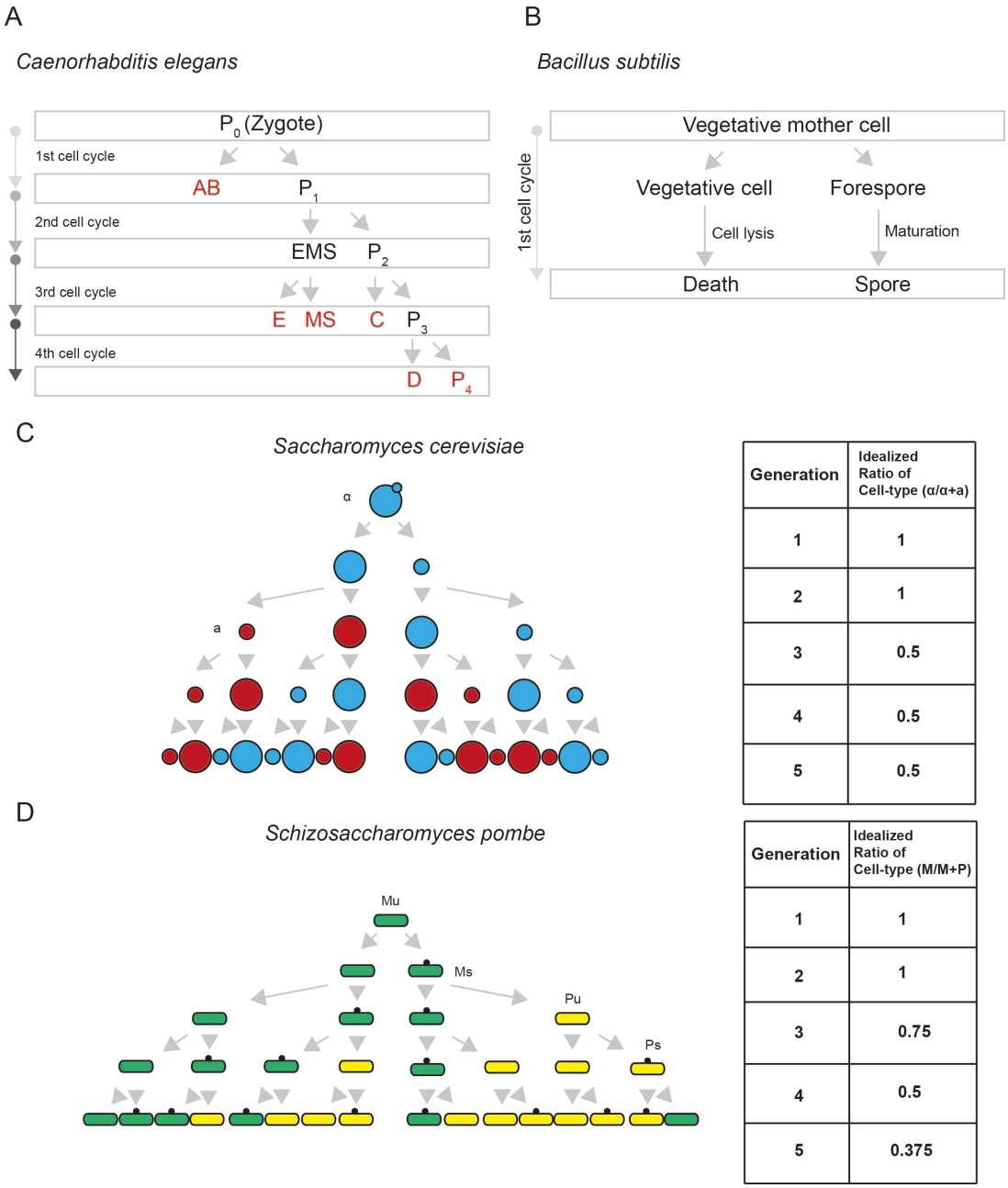
Figure 1.5. DNA-protein barriers are exploited productively by cells. In both *E. coli* **(A)** and *B. subtilis* **(B)**, terminator proteins bind to terminator sequences to terminate DNA replication within a relatively small expanse of the genome. At the *rDNA* locus, polar and non-polar replication termination prevents head-on collisions between transcription and replication. This barrier is conserved as seen in humans **(C)**, *S. cerevisiae* **(D)** and *Sz. pombe* **(E)**, although neither the cis-acting elements nor the trans-acting factors need not be completely similar.

1.7 ASYMMETRIC CELL DIVISION

Asymmetric cell division denotes a phenomenon whereby daughter cells derived **mitotically** from the same mother undergo distinct cellular fates despite having the same genotype. Such cell divisions are important for the life-cycles of both multicellular and unicellular organisms.

In *C. elegans*, upon fertilization, the zygote undergoes five asymmetric cell divisions within the first four cell cycles to produce six founder cells: AB, MS, E, C, D and P_4 (Fig 1.6) (Kipreos, 2005). These founder cells are critical in blueprinting the correct axial development of the embryo. In *B. subtilis*, environmental stimuli can trigger cells to undergo spore formation in a process that is dependent on asymmetric cell division (Fig 1.6) (Errington, 2003). Derivation of a multitude of different cell-types from toti- and pluripotent stem cells in humans is yet another example of asymmetric cell division. Meanwhile, some homothallic (self-fertile) yeasts, including *S. cerevisiae* and *Sz. pombe*, have acquired the ability to undergo asymmetric cell division to switch between cell-types (Fig 1.6). This is unique in nature as, typically, differentiated cells cannot switch to different cell-types although it is possible to artificially de-differentiate cells into induced pluripotent stem cells or to artificially reprogram differentiated cells through forced trans-differentiation (Riddle et al., 2017).

The molecular mechanism underpinning reversible cell-type switching in *Sz. pombe* and the dependence of this process on fork barriers will be explored in the next two sections of this chapter.



Generation	Idealized Ratio of Cell-type (α/α+a)
1	1
2	1
3	0.5
4	0.5
5	0.5

Generation	Idealized Ratio of Cell-type (M/M+P)
1	1
2	1
3	0.75
4	0.5
5	0.375

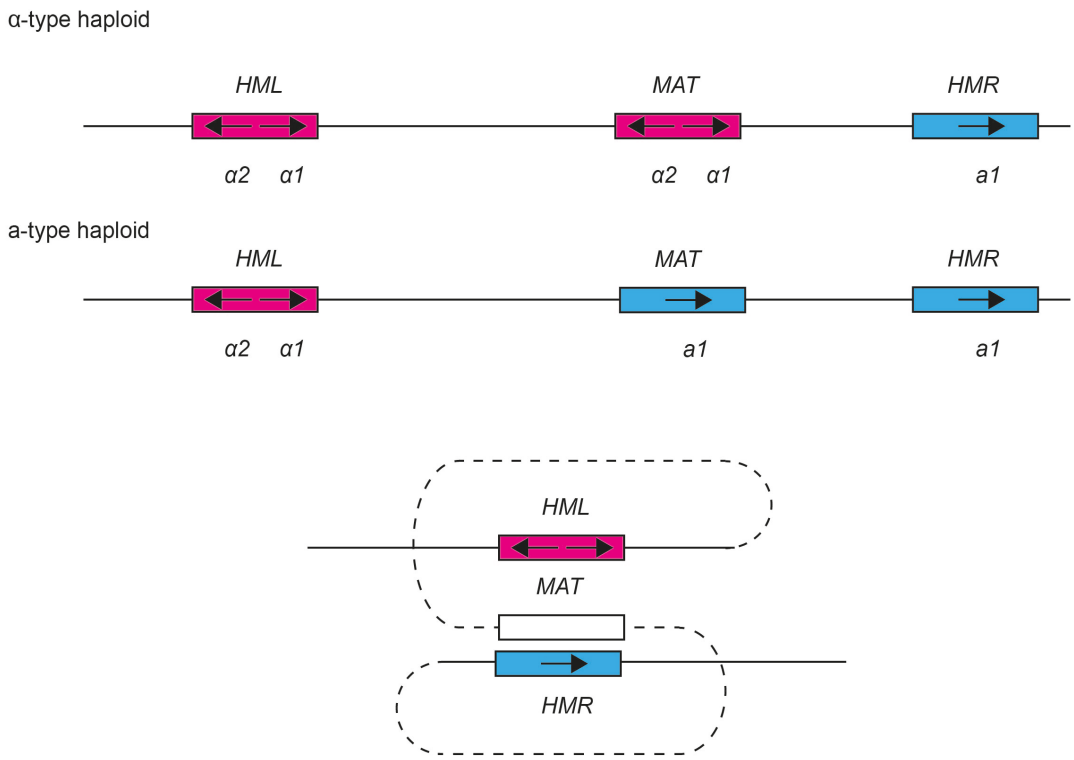
Figure 1.6. Examples of asymmetric cell type switching. (A) In *C. elegans*, asymmetric cell division gives rise to six founder cell types that are critical for development. **(B)** In *B. subtilis*, asymmetric cell division allows formation of spores. Cell-type switching in **(C)** *S. cerevisiae* and **(D)** *Sz. pombe* are also examples of asymmetric cell divisions. In this diagram, Mu and Pu cells carry *mat1M* and *mat1P* alleles respectively but are not imprinted whilst Ms and Ps cells also carry *mat1M* and *mat1P* alleles respectively and are imprinted. Imprinted cells carry a mark at the *MPS1* locus.

1.8 CELL-TYPE SWITCHING IN *SCHIZOSACCHAROMYCES POMBE*

The ability to switch cell-types allows isolated cells to recapitulate a mixed population of cells. Similar to *S. cerevisiae* (see Info Box 1), *Sz. pombe* cells spontaneously switch between h^+ and h^- cell-types and these two cell-types are sexually compatible. Under nutrient depletion, cells of opposite cell-types will mate to form a zygote that will rapidly undergo sporulation. Switching cell-types favours in-crossing by virtue of proximity suggesting that the ability to switch cell-types emerged evolutionarily not to provide genetic diversity but either as a survival tactic (spore formation) or to allow DNA repair via meiotic DNA recombination. It should be noted that *S. cerevisiae* and *Sz. pombe* have evolved from a common multicellular ancestor and acquired uni-cellularity independently from one another. Given that reversible switching between cell-types in multicellular organisms would seem to serve no purpose, the two yeast species have probably acquired the ability to switch cell-types independently (Hanson and Wolfe, 2017). Cell-type switching then might be pervasive for unicellular life-styles in eukaryotes and the underlying mechanisms could underpin our understanding of other molecular processes such as immunoglobulin class switching and, indeed, asymmetric cell division in higher eukaryotes.

The molecular mechanisms underpinning cell-type switching in *S. cerevisiae* and *Sz. pombe* are superficially similar (see Info Box 1). In *Sz. pombe*, the cell-type is determined by the identity of the gene cassette at the *mat1* locus with two transcriptionally-silenced donor loci located distally on chromosome II. These donor loci can be copied onto the *mat1* locus. Unlike budding yeast, switching is not dependent on any endonucleases, but is dependent on the configuration of the donor loci. Self-fertile (homothallic) h^{90} cells, for instance, include the *mat2P* and *mat3M* donor and are able to switch cell-type by replacing the *mat1* locus with either of the silenced *mat2P* or *mat3M* loci (Fig 1.7A). Heterothallic (non-self-fertile) strains however have different configurations at the donor loci and, hence, lose the ability to switch cell-types.

The most commonly used heterothallic strains are the h^{-S} that has lost the *mat2P* donor and the h^{+N} that has duplicated the entirety of the *mat2P-K-mat3M* region and substituted it for the *mat1* locus (Fig 1.7A). Heterothallic strains occur spontaneously in the wild, perhaps as a result of erroneous recombination events. When expressed at the *mat1* locus, the *P*- and M-cassettes express two transcripts each (P_m and P_c for the *P*-cassette and M_m and M_c for the *M*-cassette). P_c and M_c are required for conjugation between cells of opposite cell-types whilst P_m and M_m are required for meiosis within a diploid. Unlike *S. cerevisiae*, these haploid-genes when expressed at the *mat1* locus are not suppressed in diploids. Switching can occur within diploids and these can, in turn, mate to form a tetraploid or mate with a haploid to form a triploid (Egel, 1989, Gutz, 1967).



Info Box 1: Cell-type switching in budding yeast. In *S. cerevisiae*, cells switch between a and α cell-types that are sexually-compatible. This predispose cells to a sexual life cycle as sexually-compatible *S. cerevisiae* haploids readily mate to form a stable diploid that can undergo sporulation. Mating-type (or cell-type) is determined by genes expressed at the mating-type (*MAT)* locus. This locus is flanked by two (*HML* and *HMR*) transcriptionally silenced loci. At the *HML* locus, the silenced *α1* and *α2* are saved whilst, at *HMR*, the *a1* gene is housed. In a-type cells, the *MATa1* gene that encodes for the homeodomain protein a1 is expressed but removal of the *MAT* locus also pheno-copies a-type cells. By contrast, α-type cells encode the *MATα1* and *MATα2* genes that activate transcription of α-specific genes and repress transcription of a-specific genes respectively. In both haploids, haploid-specific genes are also transcribed. Switching between the two different cell-types can occur one cell cycle after germination. In cells that have divided at least once, a G_1-specific pulse of HO endonuclease creates a double stranded break (DSB) within a junction at the *MAT* locus. This break is repaired by a synthesis-dependent strand annealing (SDSA) mechanism where forks use information at the donor loci to repair the DSB. To limit futile switching, a donor bias is implemented. In a-type cells, a recombination enhancer (*RE*) sequence favours *HML* as a donor whilst, in α-type cells, the α2 product prevents *HML* to be used as a donor enabling *HMR* bias **(Haber, 2012, Hanson and Wolfe, 2017)**.

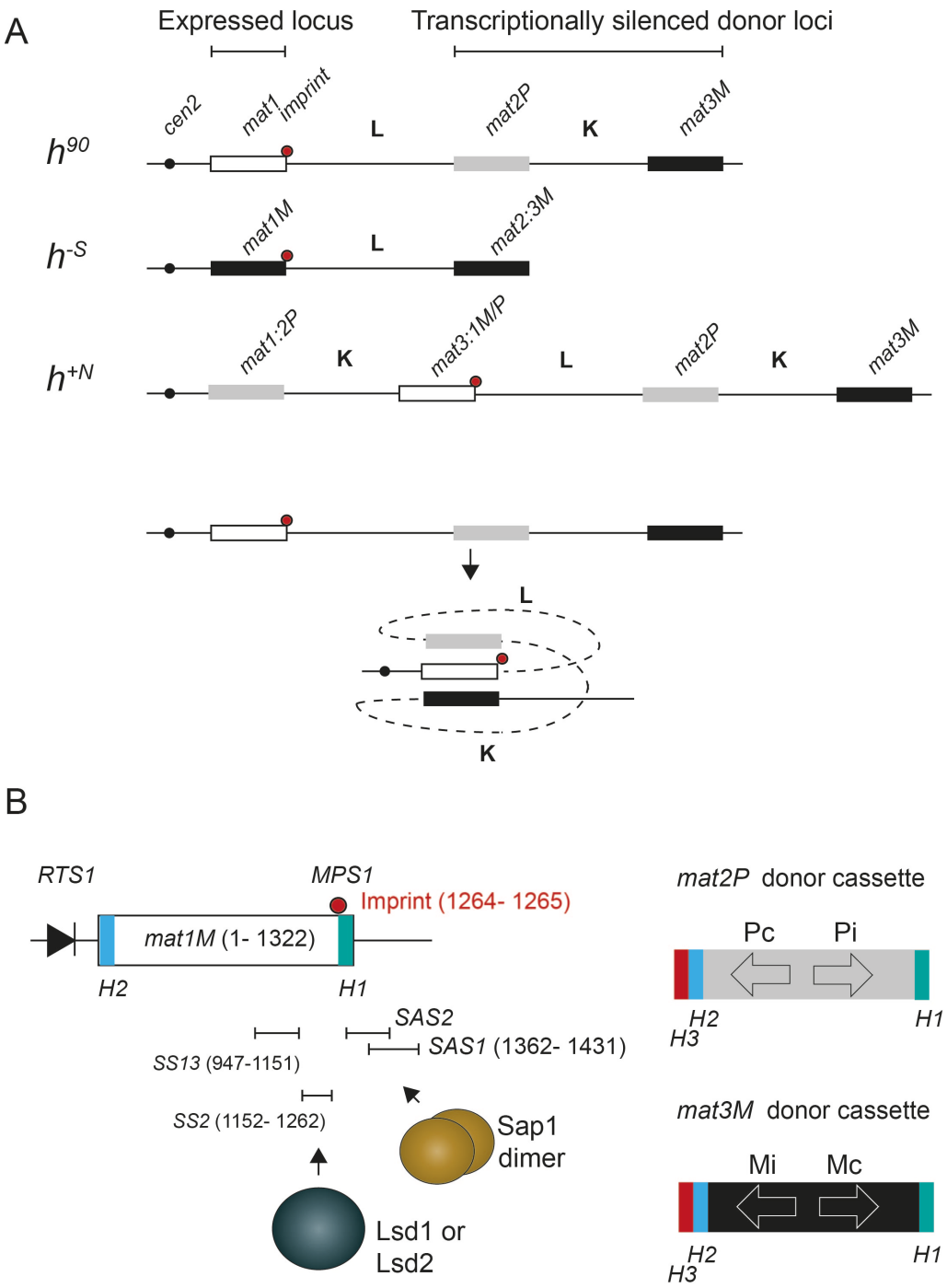


Figure 1.7. Cell-type switching in *Sz. pombe* and *cis*-acting elements requirement at the *MPS1* locus for imprinting. (A) Cell-type switching in *Sz. pombe* depends on the configuration of donor loci at *mat2* and *mat3*. Heterothallic strains cannot use their donor loci to substitute them for the gene at the expressed *mat1* locus. **(B)** Switching also depends on imprinting at *mat1*. This imprint is dependent on a number of cis-acting elements, including *SAS1*, *SAS2*, *SS2* and *SS13*. Resolution of switching requires other *cis*-acting elements, including *H1* and *H2*.

1.9 EXPLOITING THE ASYMMETRIC NATURE OF DNA REPLICATION AND POLAR FORK BARRIERS TO IMPRINT DNA AT *MPS1* AS A MARKER FOR CELL-TYPE SWITCHING

A newly germinated homothallic *Sz. pombe* cell cannot switch cell-type irrespective of the mating cassette at *mat1*. However, upon cell division one of the daughter cells inherits a mark, historically referred to as an imprint. The nature of this imprint is subject to debate: it has been reported to be either a nick (Arcangioli, 1998) or two ribonucleotides incorporated at the *mat1* pausing site 1 (*MPS1*) (Vengrova and Dalgaard, 2006).

Establishment of the imprint requires fork termination at the replication termination site 1 (*RTS1*) locus and fork pausing at *MPS1*. Both barriers require components of the FPC, $Swi1^{Tof1}$, $Swi3^{Csm3}$ and Mrc1 for optimal barrier activity (Dalgaard and Klar, 2000, Dalgaard and Klar, 2001, Zech et al., 2015). Pausing at *MPS1* is necessary but not sufficient for imprinting and the two processes have been isolated genetically (Dalgaard and Klar, 2000, Sayrac et al., 2011). In wildtype cells, the imprint is degraded during purification of genomic DNA, leading to a double stranded break at *mat1*. Consequently, Southern blotting of a HindIII fragment that includes this imprint leads to three distinct DNA fragments: the intact fragment and two fragments that arise because of breakage. By contrast, strains that are defective for imprinting do not undergo breakage at *mat1* and only the intact HindIII fragment is observable in Southern blots (Singh and Klar, 1993). Through Southern blotting of the HindIII fragment at *mat1*, it was shown that a mutant of *pol1*, *swi7-1* (*pol1G116E*), is defective for imprinting whilst not affecting pausing at *MPS1* on 2D gels (Singh and Klar, 1993, Dalgaard and Klar, 2000). Given that Pol α also contains primase subunits ($Spp1^{Pri1}$ and $Spp2^{Pri2}$), it is thus tempting to suggest that the imprint is indeed RNA in nature. If this is the case, it would suggest that there is a protective mechanism that prevents displacement of the ribonucleotides at *MPS1*, especially by Pol δ. This putative protective mechanism may take the form of heterochromatin that is known to impede DNA synthesis by Pol δ that otherwise synthesizes DNA continuously,

in an uninterrupted fashion (Devbhandari et al., 2017). In fact, binding of Sap1 (also involved at the *rDNA RFB1*) and of a histone demethylase, Lsd1 within the vicinity of the imprint are required for stable imprinting at *MPS1* (Arcangioli and Klar, 1991, Holmes et al., 2012) (Fig 1.7B). These may contribute to a chromatin architecture that is favourable for protecting ribonucleotides at *MPS1*. Beyond its effects on cell-type switching and on the S phase response to alkylation damage (Koulintchenko et al., 2012), the *swi7-1* allele is ill-characterized. The allele is also known to confer a mutator phenotype and is characterized by hypersensitivity to both HU and MMS. However, the strength of interaction between the Swi7-1 protein with different proteins such as Mcl1 or other subunits of Pol α (with the exception of Swi6) and how it differs from its wildtype counterpart has not been determined.

Once the cells are imprinted, the imprint itself is stabilized for one cell cycle. In the subsequent S phase, the imprint at *MPS1* blocks leading strand replication. Given that forks are still terminated at the *RTS1* locus, the entirety of *mat1* locus and its immediate surrounding cannot be replicated conventionally (Dalgaard and Klar, 2001). Instead, forks are reversed at *MPS1* and a 'chicken foot' intermediate is observable on 2-D gels. This intermediate is resolved by the RecQ-like helicase Rqh1 (Vengrova and Dalgaard, 2004).

Fork reversal seems to precede fork invasion of either the *mat2P* or *mat3M* loci. Homology is provided by the *H1* and *H2* sequences that allow intra-chromosomal recombination between the newly copied donor and the cassette at *mat1*. Sequences not found between the *H2* and *H1* sites seem to be digested. This would explain why the *mat1* locus does not contain the *H3* site although both donor loci do. This synthesis-dependent strand annealing (SDSA) mechanism was verified *in vivo* by using artificial *mat2P* and *mat3P* alleles (Yamada-Inagawa et al., 2007). Resolution of the recombination between the *mat1* and its donor loci is resolved by concerted action of $Msh2^{Swi8}$, $Msh3^{Swi4}$, $Rad16^{Swi9}$ and Swi10 (Arcangioli and de Lahondès, 2000). Puzzlingly, absence of donor loci does not lead to cell death suggesting that the lack of replication between *RTS1* and *MPS1* can be compensated for by inter-chromosomal repair using the intact sister chromatid as a template (Klar

and Miglio, 1986). The various steps involved in cassette-switching at *mat1* are summarized in Fig 1.8.

It should be noted that mutants that are defective in cell-type switching are not necessarily defective for imprinting (Gutz and Schmidt, 1985). Whilst switching-defective mutants can be distinguished by iodine staining, imprinting-defective mutants are characterized by non-breakage of the *mat1* cassette during purification of genomic DNA.

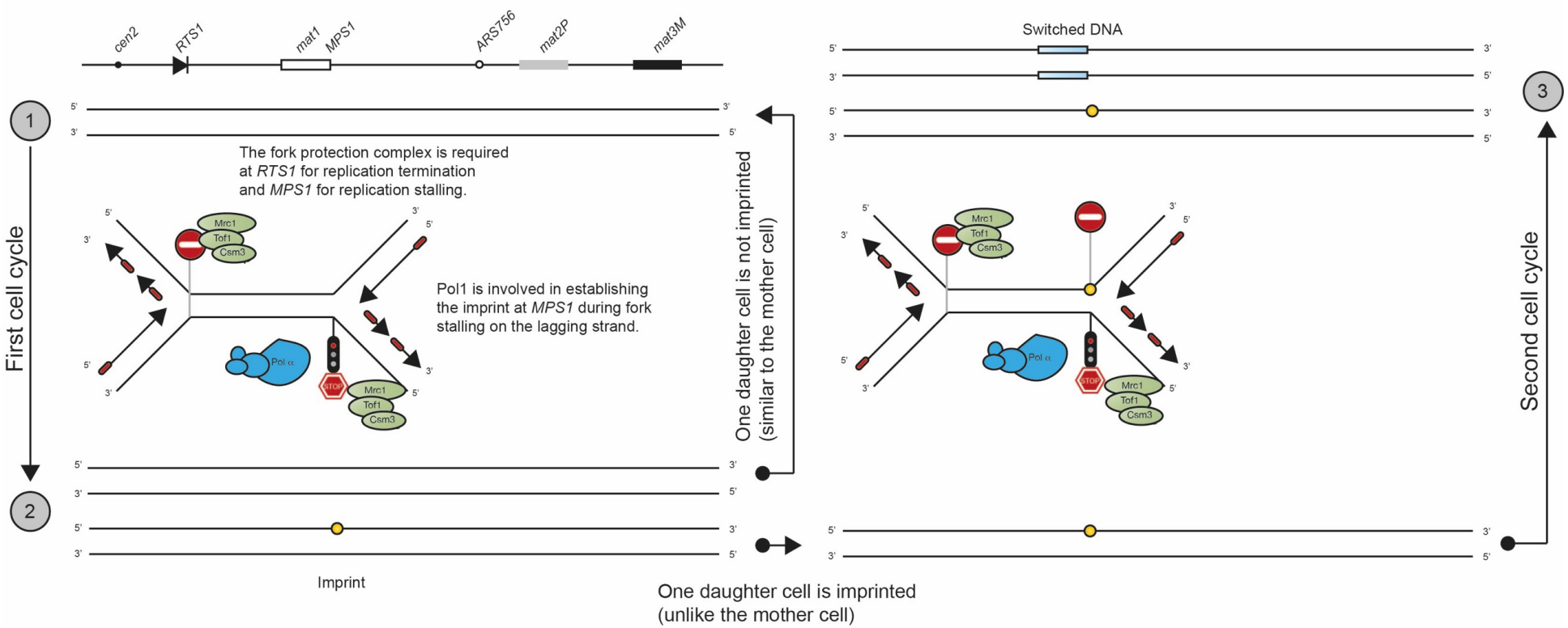


Figure 1.8. Cell-type switching in *Sz. pombe* requires two fork barriers. In the first cell cycle, forks approaching proximally (from *cen2*) are terminated at *RTS1* and forks approaching distally are paused at *MPS1*. Pausing and termination both require Swi1, Swi3 and Mrc1. The pausing event is required for Pol1-dependent imprinting at *MPS1*. In the following cell cycle, the imprint act as a second terminator site so that the DNA between *RTS1* and *MPS1* is not replicated conventionally. Instead, it needs to be copied from information stored at either one of two donor loci (*mat2P* or *mat3M*). This leads to cassette switching.

Like in *S. cerevisiae* (see Info Box 1), futile switching of cell-types is prevented by means of donor preference. Indeed, from a newly germinated cell, one of four the grand-daughter cells has an about 90% chance of switching cell-type (Miyata and Miyata, 1981). Donor preference is mediated by both *cis*- and *trans* acting-sequences (Thon and Klar, 1993, Jakočiūnas et al., 2013). In cells carrying the M-cassette at *mat1* (*mat1M*), donor preference for *mat2P* is mediated by relatively high concentrations of Swi2 (and its partner, Swi5) and by the *SRE2* enhancer region found distally to *mat2P*. Conversely, in cells expressing the *P*-cassette at the *mat1* locus (*mat1P*), *mat3M* donor bias is mediated by the enhancer *SRE3* found distally to *mat3M* and by relatively lower expression of *swi2* and *swi5* (Jakočiūnas et al., 2013).

Importantly, the donor loci at *mat2* and *mat3* are silenced by virtue of a combination of *cis*-acting elements found proximal to the donor loci (Thon et al., 1994, Thon et al., 1999) and *trans*-acting factors, among which the HP1-like chromodomain protein Swi6 (Nakayama et al., 2000) and the H3K9 methyltransferase Clr4 (Nakayama et al., 2001b). Swi6 has been shown to interact physically with both $Dfp1^{Dbf4}$ (Hayashi et al., 2009) and $Cdc18^{Cdc6}$ (Li et al., 2011), modulating replication at heterochromatin-rich sites in the *Sz. pombe* genome. The *mat2* and *mat3* loci are replicated early in S phase (Kim et al., 2003). Both the early replication and lack of transcription at these loci could be a result of the compactness of the DNA there.

Intriguingly, switching-defective alleles of *swi1* (*swi1-111*), *swi3* (*swi3-146*) and *pol1* (*swi7-1*) were found to de-suppress silencing at the donor *mat2* and *mat3* loci. Higher levels of de-suppression were observed with the *swi7-1* allele (Nakayama et al., 2001a). It was determined that recombinant Pol1 interacts with Swi6 by virtue of its C-terminus (residues 1032-1405) *in vitro* and that endogenous Pol1, but not Swi7-1, interacts with Swi6 *ex vivo*. Moreover, *swi7-1* strains were characterized by reduced levels of Swi6 at the silenced donor loci, centromeres and telomeres (Nakayama et al., 2001a). This indicates that Swi6 is enriched at heterochromatin-rich DNA in a Pol1-dependent manner and that Swi7-1 cause de-suppression of the silenced donor loci by virtue of reduced interaction with Swi6. A similar silencing defect

was seen in strains harbouring a temperature-sensitive of *pol1* with checkpoint defects (Murakami and Okayama, 1995), *pol1-H4* (*pol1G889D*) (Ahmed et al., 2001). Similar to the Swi7-1, Pol1-H4 has reduced interaction with Swi6 and Swi6 levels at the silent donor loci are much reduced in strains harbouring the *pol-H4* allele. Interestingly, both the temperature-sensitive and cycle defects of *pol1-H4*, but not its silencing defect, were suppressed by co-expression of either wildtype *pol1* or *cds1* (Ahmed et al., 2001, Murakami and Okayama, 1995). This would suggest that the silencing defect was dominant, perhaps reflecting a gain of function. The *pol1-H4* mutant, as well as two other temperature-sensitive mutants, *pol1-ts11* and *pol1-ts13*, show observable silencing defects when crossed in strains carrying the heterothallic *mat1-Msmt0* mutation (Ahmed et al., 2001). Strains carrying the *swi7*-1 and *mat1-Msmt0* mutations did not present these defects unless also deleted for a 1.5 kb *BglII-BssHII* fragment proximal to *mat2P*. As such, it is tempting to suggest that the overlapping phenotypes between *swi7-1* and *pol1-H4* might be underpinned by un-identical mechanisms at the molecular level.

Consequently, the exact role of Polymerase α in the mating-type switching of fission yeast, especially in imprinting at the *MPS1* locus is still unknown. Elucidating this problem could inform us about the role of DNA polymerases in response to pre-programmed fork barriers, perhaps in relation to fork restart after stalling. This will be the focus of Chapter 3 of this thesis.

1.10 THE TRANSCRIPTION PROCESS AS A BARRIER TO DNA REPLICATION

Both DNA replication and transcription are fundamental processes for life and represent central pillars of molecular biology. Whilst DNA replication is restricted to S phase, transcription can occur throughout the cell cycle. Indeed, in human cells, transcription of longer genes can span the better part of a cell cycle or can even last longer than one cell cycle. For instance, transcription of the human dystrophin gene (*DMD*) lasts around 16 h (Tennyson et al., 1995) whilst transcription of the longest human gene, *CNTNAP2* (functional in the nervous system) takes between 9 to 35 h to complete (Helmrich et al., 2011). On the other hand, the fast-growing HeLa cells have a doubling time of about 23 h. DNA replication and transcription then can overlap temporally. Moreover, both processes utilize the same substrate (the double helix) to which they impose transient but significant architectural alterations. Such altered DNA are not ideal substrates for either DNA replication or transcription, and can halt their respective progression. Importantly, collisions between DNA replication and transcription, especially head-on collisions, lead to genomic instability (Fig 1.9) (Liu and Alberts, 1995, Helmrich et al., 2011, Helmrich et al., 2013, Prado and Aguilera, 2005).

In prokaryotes, forks originate from a single origin. This neatly halves bacterial chromosomes in two replichores. It has been established that the direction of transcription of operons on either replichore are biased so that replication forks and transcription complexes encounter in a mostly co-directional fashion. This suggests an evolutionary advantage to reducing the frequency of head-on collisions between the two processes. For example, in *E. coli*, the genome is organized so that 55% of protein coding genes align with the direction of replication whilst all seven of the ribosomal RNA operons and 53 out of 86 tRNA genes (62%) also align with the direction of forks (Blattner et al., 1997). Operons of highly transcribed genes are also more likely to be orientated so that the 5’ end of the genes occur proximal to the origin (Brewer, 1988). The bias is starker in *B. subtilis* where the genome is organized in such a way that

75% of the predicted genes are transcribed co-directionally with DNA replication (Kunst et al., 1997). Beyond genome organization, there are several lines of evidence that demonstrate that the transcription process represents a potent barrier to fork progression. For example, in asynchronous cultures of *E. coli*, rifampicin treatment leads to increased replication speeds (Pato, 1975). Rifampicin works by preventing synthesis of novel transcripts but does not interfere with the elongation of nascent mRNA molecules. Thus a reduction in the number of transcribing complexes correlates with increased fork speeds. In the same study, the author also found that streptolydigin treatment leads to reduced fork speeds (Pato, 1975). Streptolydigin also affects transcription but by slowing down the process (Brewer, 1988). As such, stabilizing transcribing complexes on the genome correlates with reduced fork speeds. This fork barrier activity of the transcription process can be reproduced *in vitro* (Liu and Alberts, 1995). Interestingly, replication pauses for longer when the two processes undergo head-on rather than co-directional collisions (Liu and Alberts, 1995).

Head-on collisions between transcription and DNA replication have also been shown to be more deleterious than co-directional collisions in eukaryotes. Using a plasmid-based assay, Prado and Aguilera have shown that head-on collisions between replication forks and transcribing complexes lead to increased levels of recombination compared to co-directional collisions in *S. cerevisiae* (Prado and Aguilera, 2005). The authors also showed that the extent of recombination was highest when the promoter of the transcription module was S phase specific (*HHF1* promoter). Similar studies have been carried out at genomic loci. One notable example involved using isogenic yeast strains carrying the *GAL1-lys2*^{frameshift} construct in the presence or absence of Gal80 (Datta and Jinks-Robertson, 1995). The *GAL80* gene supresses transcription from the *GAL1* promoter in the absence of galactose so that strains carrying the *gal80*Δ allele constitutively express from the *GAL1* promoter. The cells were grown in medium lacking galactose. Cells with constitutive expression of the *GAL1* promoter were more likely to undergo *LYS2* reversion demonstrating that transcriptionally active DNA is characterized by increased recombination, perhaps by hindering other

processes such as DNA replication (Datta and Jinks-Robertson, 1995). In a similar study where the *lys2*^{frameshift} mutant was expressed under the control of a repressible promoter, it was determined that the extent of *LYS2* reversion was directly proportional to the levels of transcription (Kim et al., 2007). The authors also found that mutational signature was co-dependent on the orientation of the *lys2*^{frameshift} allele relative to closest origin from which it was cloned (*ARS306*) and its expression level. When the 5' end of the gene was found distal to the origin, large deletions were more likely at low transcription levels whilst complex indels (as well as other mutants) were more prevalent at high transcription levels. As such, conflicts between DNA replication and transcription can lead to deleterious recombogenic events, distinct from those arising when forks do not meet transcribing complexes.

It is evident that the transcription process need not impede DNA replication directly. Indeed, it has been suggested that supercoils accumulating ahead of the replication fork and transcription machinery might provide a topological constrain that leads to fork reversal and genomic instability (Helmrich et al., 2013). Moreover, the transcription process generates an intermediate known as the R-loop that impedes replication forks. The following chapter will be devoted to R-loops and the factors that contribute to their formation and stability within genomes.

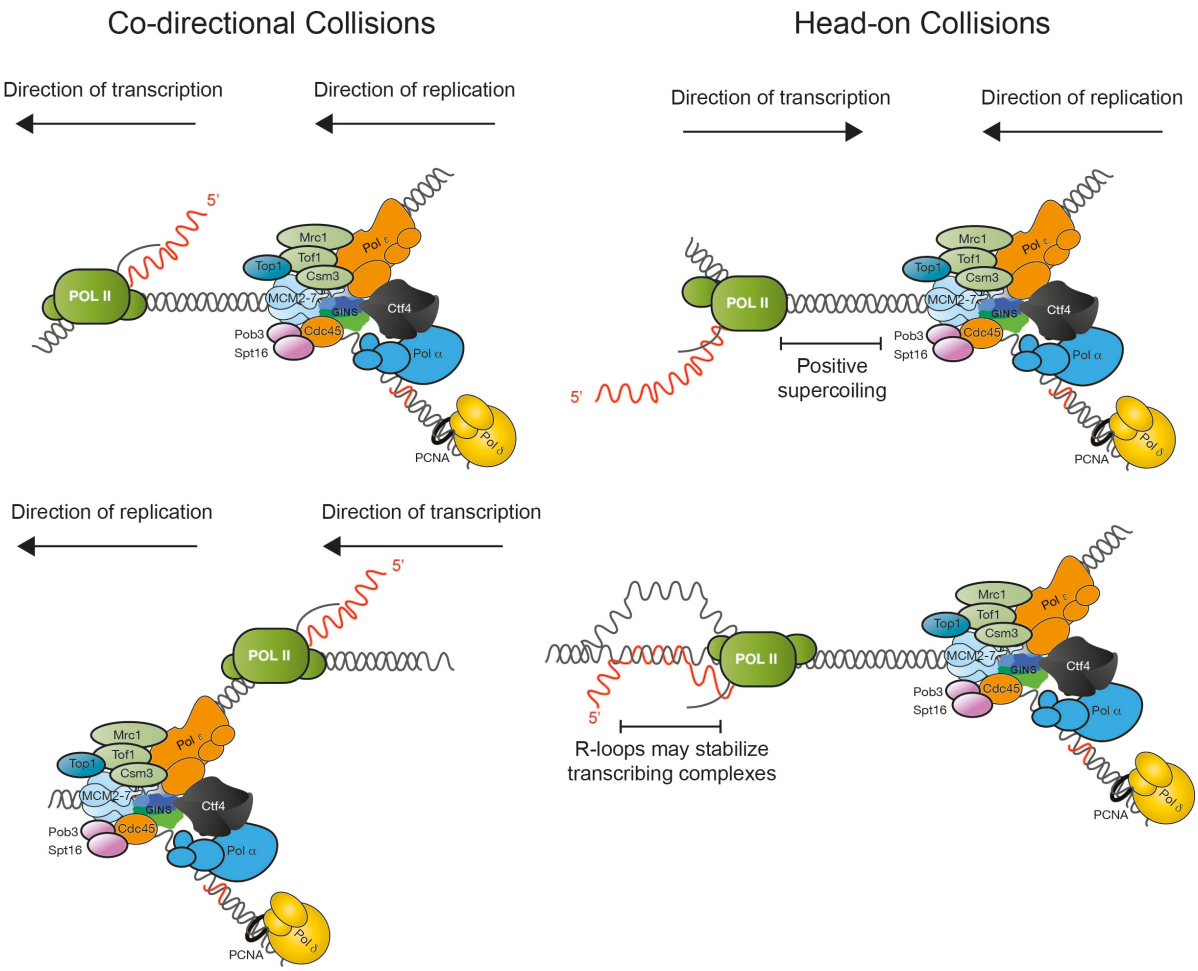


Figure 1.9. DNA replication and transcription can interfere with the progression of one another. Collisions between the two processes can occur co-directionally. In eukaryotes, the two processes have comparable speeds, but either forks or transcribing complexes may pause. This creates the opportunity for co-directional collisions. Head-on collisions may also happen if forks travel towards transcribing complexes. The two types of collisions trigger different DNA damage responses and appear to resolved differently (Hamperl et al., 2017). Importantly, head-on collisions are more harmful for cells (Prado and Aguilera, 2005).

1.11 R-LOOP BIOLOGY

The R-loop is a three-stranded nucleic species made up of both DNA and RNA. This hybrid consists of a nascent transcript reannealing to its template DNA, with the non-template strand displaced as ssDNA (Fig 1.10) (Mirkin and Mirkin, 2007, Aguilera and García-Muse, 2012).

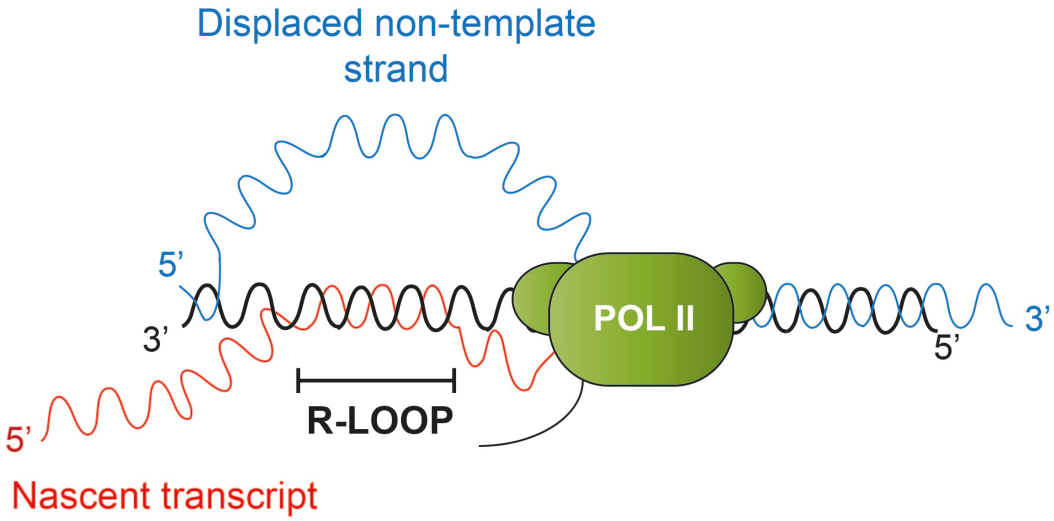


Figure 1.10. The R-loop is a three-stranded DNA/RNA hybrid that arises as an intermediate of the transcription process. It is made up of a nascent RNA molecule bound to its template, with the non-template strand displaced.

Importantly, R-loops are thermodynamically more stable than their equivalent dsDNA by virtue of the hybrid interaction between RNA and DNA molecules (Lesnik and Freier, 1995, Roberts and Crothers, 1992) and need to be removed enzymatically. Structural studies suggest that R-loops adopt a conformation that is intermediate to that of A- and B-form DNA (Shaw and Arya, 2008). This may have implications for the homeostasis of the nucleic species and could underpin its nature as a fork barrier. Whilst it has been established that R-loops arise as by-products of the transcription process, the precise mechanism through which they are formed is unknown. It is thought that R-loops are formed through a 'thread-back' model where the nascent mRNA molecule reanneals to its template strand once it exists its RNA polymerase (Fig 1.11). This would fit crystallographic data that shows that the template strand and nascent mRNA are extruded through different passages

within eukaryotic RNA polymerases (Westover et al., 2004). Moreover, R-loops synthesized *in vitro* using the prokaryotic T7 RNA polymerase are susceptible to RNase T1 (an enzyme that degrades single-stranded RNA at G-residues) added during transcription but not after completion of the transcription process, consistent with the thread-back model (Roy et al., 2008). However, it is not impossible that, under some conditions, the nascent transcript might not disengage from its template but instead exit the polymerase as a hybrid, promoting R-loop formation (Fig 1.11).

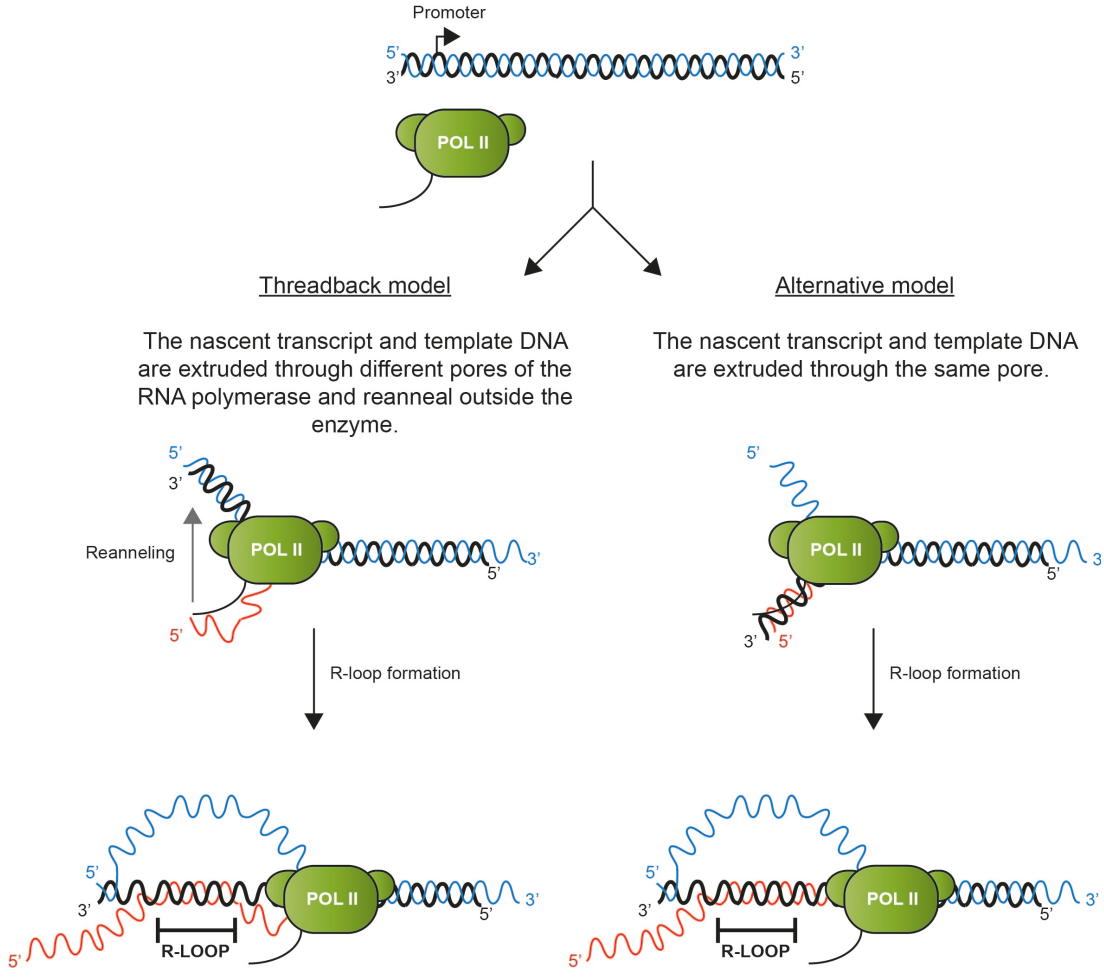


Figure 1.11. Two models have been proposed for the formation of R-loops. In the thread-back model, the nascent transcript and template strand are extruded via different channels. However, the transcript then reanneals with the template strand, displacing the non-template strand. Empirical data favours this model. An alternative model suggests that the nascent transcript emerges bound to its template from the RNA polymerase, promoting R-loop formation.

R-loop formation is favoured by a number of factors, as reviewed in (Aguilera and García-Muse, 2012) (Fig 1.12). Firstly, the nature of the DNA needs to be favourable. Indeed, GC skew (or strand asymmetry) at either the 5' end of genes (immediately downstream of the promoter) or at their 3' end has been found to promote R-loop formation (Ginno et al., 2012, Ginno et al., 2013). The strength of the promoter also correlates positively with R-loop formation (Huertas and Aguilera, 2003, Hamperl et al., 2017). Importantly, the replication process also plays a critical part in R-loop homeostasis. In a recent paper from the Cimprich lab, it was shown that head-on, but not co-directional, collisions between replication forks and transcribing complexes favoured the formation of R-loops (Hamperl et al., 2017). In fact, head-on and co-directional collisions between the replication fork and transcribing complexes were found to lead to distinct types of genomic instability that are also resolved through distinct pathways (Hamperl et al., 2017). Head-on collisions on substrates prone to form R-loop were found to trigger robust phosphorylation of γH2AX and ATR^{S428} as well as phosphorylation of ATR targets $RPA32^{S33}$ (a component of the human replication protein A), and $Chk1^{S345}$ suggesting that head-on collisions trigger and are resolved by the ATR pathway in human cells. Conversely, co-directional collisions were seen to lead to robust auto-phosphorylation of ATM^{S1981} as well as phosphorylation of ATM targets $KAP1^{S824}$, $Chk2^{T68}$ and $RPA32^{S4/8}$ (Hamperl et al., 2017). This would provide a rationale on how head-on collisions would favour formation of R-loops whilst co-directional collisions would lead to R-loop removal. Interestingly, substituting the substrate for one that is not susceptible for R-loop formation does not lead to ATR and ATM activation upon head-on and co-directional collisions respectively (Hamperl et al., 2017). Taken together, this would suggest that spontaneously-arising R-loops trigger genomic instability upon collisions between the replication and transcription process and that, in the case of head-on collisions, this leads to accumulation of R-loops, potentially forming a positive feedback that would lead to profound genomic instability.

1.12 THE PHYSIOLOGICAL IMPORTANCE OF R-LOOPS

As illustrated above, R-loops occur naturally. Yet they may be harmful for genome stability and integrity (Aguilera and García-Muse, 2012). It is intriguing then why cells have not evolved to completely prevent R-loop formation in their genomes. In actuality, several lines of evidence suggest that R-loops are physiologically useful in several pathways, including in immunoglobulin class switching in B-cells and in DNA replication.

In E. coli, in vivo replication of ColE1 plasmids has been shown to be dependent on both the primase *dnaA* and polymerase *polA*. Remarkably, replication of ColE1 plasmids was inhibited by treatment with the transcription-inhibitor rifampicin *in vivo* (Tomizawa, 1975). Likewise, i*n vitro* replication of ColE1 plasmids (unlike that of chromosomal DNA) was also inhibited by rifampicin and was dependent on the presence of rNTPs (Sakakibara and Tomizawa, 1974). These observations strongly suggested that transcription is also necessary for DNA replication of ColE1 plasmids. *In vitro* assays were also used to demonstrate that a transcription initiation site found some 550 base pairs upstream of the ColE1 origin of replication produces an R-loop that is processed by the RNase H1 endonuclease (*rnhA*) to produce a 3'-hydroxyl end (Itoh and Tomizawa, 1980). The R-loop, then, is effectively processed into an RNA primer that can be used to initiate the replication process. Interestingly, whilst *dnaA* is an essential gene in *E. coli* (Baba et al., 2006), inactivation of either *dnaA* or of the chromosomal origin of replication, *oriC*, is not lethal in *rnhA* mutants, presumably by virtue of increased stability of R-loops (Kogoma and von Meyenburg, 1983, Kogoma, 1997). This suggests that, at least in prokaryotes, loss of priming on chromosomal DNA can be compensated for by the presence of R-loops.

Replication of mitochondrial DNA also requires R-loops. This is true for the circular human mitochondrial DNA where transcripts lead to the formation of RNAse H-sensitive R-loops whose 3' ends map to the origins of replication in the mitochondrion (Xu and Clayton, 1996). Electron microscopy data also

suggests that stable R-loops are intermediates of the replication process in human as well as other mammalian mitochondria (Pohjoismäki et al., 2010). More surprisingly, R-loops were also found to be important for replication of the linear mitochondrial genome in *S. cerevisiae*. Yeast mtDNA has several origins (seven to eight) (Foury et al., 1998). Work focusing on two of those origins (*ori1* and *ori5*) found that, *Aspergillus* nuclease S1-insensitive but RNase-sensitive nucleic species were required for DNA replication initiation at those loci (Baldacci et al., 1984). The S1 nuclease digests single-stranded nucleic acid species (both DNA and RNA), suggesting that R-loops participate in DNA replication at mitochondrial origins in yeast despite yeast mtDNA being linear. As such, R-loops are not precluded from initiating replication in linear genomes and could participate in the replication of eukaryotic genomes.

R-loops are involved in gene expression. In *Arabidopsis*, differences in flowering phenotypes can be attributed to the relative expression and suppression of the floral repressor gene, *FLOWERING LOCUS C* (FLC) that is receptive to changes in temperature. Several pathways contribute to changes in the levels of *FLC* transcription but one involves stabilizing R-loops whereby these then compete with long non-coding RNA (lncRNA) at the *FLC* locus. LncRNA contribute to the epigenetic silencing of the *FLC* locus so that R-loops effectively act as positive regulators of *FLC* expression in *Arabidopsis* (Csorba et al., 2014, Sun et al., 2013). R-loops might also regulate gene expression by mediating changes in chromatin. Indeed, work in the Aguilera lab found that increasing the levels of R-loops by using deletion mutants of *RNH1* and R*NH201* or *HPR1* led to a concomitant increase in the levels of the heterochromatin marker H3S10 in *S. cerevisiae* along the lengths of most ORFs (Castellano-Pozo et al., 2013). The enrichment of the H3S10 marker was most marked at centromeric and pericentromeric loci (Castellano-Pozo et al., 2013). Counterintuitively, H3S10 also acts as a marker for highly transcribed regions at euchromatin (Sawicka and Seiser, 2012). As such, dysregulation of R-loops leads to global changes in the accessibility of DNA and, consequently, to altered transcription profiles globally in the cell. In the same study, the authors also showed that elevating levels of R-loops in the nematode *C. elegans* and in HeLa cells led to the enrichment of the H3S10

marker, suggesting that the mechanism linking R-loop to chromatin rearrangement is conserved across eukaryotes. Meanwhile, R-loops arising from antisense transcription were found to mediate the formation of dsRNA in human as well as in murine cells (Skourti-Stathaki et al., 2014). dsRNA is an intermediate in the recruitment of the H3K9me2 chromatin marker. As such, some R-loops affect gene expression through chromatin silencing. By contrast, other R-loops inhibit the methylation (and silencing) of the CpG island promoters of several genes (Ginno et al., 2012).

R-loops also contribute to gene expression through transcription termination (Ginno et al., 2013). In fact, failure to remove R-loops at the 3'-end of genes leads to termination read-through of transcripts (Mischo et al., 2011, Skourti-Stathaki et al., 2011). However, in human cells, complete removal of R-loops by overexpressing RNase H1 also leads to transcription termination defects (Skourti-Stathaki et al., 2011). These apparently contradictory results could be reconciled by envisaging a threshold of R-loops that is required for transcription termination in human cells but enrichment beyond a certain level can be detrimental to the cell. An observation apparently in support of such a model is that the overexpression of RNase H1 leads to synthetic defects, presumably by removing R-loops below a certain physiologically beneficial threshold (Paulsen et al., 2009).

R-loops are also implicated in telomere dynamics. Telomeres cannot be replicated and are progressively shortened in each successive round of DNA synthesis, unless acted upon by telomerase, a reverse transcriptase. This leads to replicative senescence and has been linked to the both the Hayflick limit and ageing (Bekaert et al., 2005, Olovnikov, 1996). Remarkably, telomeric length can be preserved in spite of compromised telomerase activity, such as in a Rad52-dependent manner in *S. cerevisiae* (Lundblad and Blackburn, 1993). R-loops- in the form of TERRAs (telomeric-repeat-containing RNAs)- naturally form at telomeres. The abundance of these R-loops increase upon deletion of components of the THO/TREX complex that are involved in mRNA processing (including *HPR1* and *THO2*) (Pfeiffer et al., 2013) or of the RNase H enzymes (*RNH1* and *RNH201*) (Balk et al., 2013)

and upon shifting strains encoding the temperature-sensitive *rat1-1* (Rat1 is involved in transcription termination) to non-permissive temperatures (Luke et al., 2008). R-loops at telomeres trigger recombination, presumably by acting as fork barriers. The resolution of this recombination is important for the maintenance of telomeric length in wildtype cells but would seem to accelerate telomere shortening and cellular senescence in recombination-compromised cells (Balk et al., 2013). As such, R-loops are involved in telomerase-independent telomere maintenance. In fact, in telomerase-deficient cells, R-loops promote recombination preferentially at critically shortened telomeres (Graf et al., 2017). It should also be noted that, in strains defective for both telomerase and recombination, *rad5*Δ causes additional defects (Fallet et al., 2014). Rad5 is part of the error-free DNA-damage tolerance pathway. It is an E3 ubiquitin ligase and also possesses DNA-dependent ATPase activity (Xu et al., 2015). Both enzymatic activities are necessary for error-free DNA-damage tolerance. Rad5 is recruited at telomeres, enabling error-free repair at those loci. Given that *rad5*Δ causes synthetic defects in strains defective for both telomerase and recombination, it is possible that R-loops are involved in the Rad5-dependent repair of telomeres.

Given the role of R-loops in the Rad52-driven recombination at telomeres, it is not surprisingly to note that R-loops promote recombination at other loci in the genome. This was illustrated in an elegant study in the Storici lab where yeast cells were constructed to contain two chimeric *HIS3* genes. At the endogenous locus on chromosome XV, the *HIS3* gene was disrupted with a site that can be specifically digested to a double-stranded break by the HO endonuclease whilst a second *HIS3* allele was introduced at an ectopic site on chromosome III. This second *HIS3* allele was disrupted with an artificial intron and under the control of the *GAL1* promoter so that an antisense his3 transcript can be used to repair the endogenous *HIS3* gene after having undergone splicing. Alternatively, a single chimeric *HIS3* allele was used on chromosome III where the *HO* site was introduced in the artificial intron, still under control of the *GAL1* promoter so as to generate an antisense transcript. In this case, the transcript was tested for the ability to self-heal the DNA from which it was transcribed (Keskin et al., 2014). Histidine prototrophs were

recovered in wildtype but not in *spt3*Δ cells. Removal of Spt3 leads to a reduction of transcription of *Ty* transposons that could reverse-transcribe the antisense spliced his3 transcript to *HIS3* cDNA. The latter could then be used instead of the transcript for recombination. As such, in wildtype strains, recombinants were exclusively a result of the presence of *HIS3* cDNA. However, histidine prototrophs were obtained in strains triply deleted for *SPT3*, *RNH1* and *RNH201*, suggesting that R-loops can lead to repair of double-stranded break directly without having to recourse to reverse-transcription first (Keskin et al., 2014). This repair pathway was also shown to be dependent on Rad52 but not Rad51.

Finally, R-loops are important for immunoglobulin class-switching in activated B-cells, leading to the synthesis of immunoglobulins other than IgM (Chaudhuri and Alt, 2004, Yu and Lieber, 2003). Class-switching involves an atypical recombination named class switch recombination (CSR) that eschews recombination based on homology between different DNA loci and instead makes use of DSBs to excise portions of the genome that encode for antibody heavy chains. After excision, the DNA between the DSBs assembles itself in a promoter-less circular DNA and is presumably lost by the cell. The remaining DNA is repaired by non-homologous end joining (NHEJ). Inversion of the donor $S_{\gamma}1$ locus in mice, promoting anti-parallel transcription, was found to severely impair isotype switching to IgG1 (Shinkura et al., 2003), suggesting that transcription-related intermediates are important for CSR. Cell-free transcription of the donor S_{μ}, $S_{\gamma}2b$ and $S_{\gamma}3$ loci led to nucleic species that migrated slowly on agarose gels and that were RNase A-insensitive but RNase H-sensitive (Daniels and Lieber, 1995). In fact, the nature of the DNA at class switch donor sequences was found to be ideal for R-loop formation (Roy et al., 2008). Although the precise mechanism through which R-loops are used for CSR is yet to be established, a seductive hypothesis is that R-loops could stabilize the non-template strand of DNA, the latter being a substrate for activation-induced cytidine deamination where cytidines are deaminated to uridines. Uridines occurring on the non-template strand would then be prone to digestion, promoting abasic DNA and eventually DSBs.

1.13 R-LOOPS AS A SOURCE OF GENOME INSTABILITY AND THE MECHANISMS OF R-LOOP REMOVAL IN EUKARYOTES

R-loops were previously considered to be rare by-products and transient intermediates of the transcription process. Consequently, the impact of R-loops, beyond their very specific uses (as illustrated in Section 1.12) on the genome was deemed negligible. It is quickly becoming apparent however that R-loops are generated much more frequently than previously envisaged (Aguilera and García-Muse, 2012). Given also that R-loops are more stable than their dsDNA counterpart, that the displaced non-template DNA is more susceptible to damage and that they can plausibly interfere both with fork progression and the transcription process, R-loops are increasingly being appreciated for the threat they pose to both genomic integrity and to gene expression.

In theory, there are several ways through which R-loops could contribute to genome instability. By virtue of their structural nature alone, the non-template strand in R-loops are exposed as ssDNA and prone to chemical modifications. Some of these such as activation-induced deamination in B-cells could be programmed and beneficial whilst other modifications could contribute to unwanted transcription-associated mutagenesis (TAM). Indeed, C-G to T-A transition mutations occur 140 times more frequently in ssDNA than in dsDNA (Frederico et al., 1990) whilst, in *E. coli,* C-G to T-A mutations that trigger reversion of a reporter kanamycin gene are favoured when the target cytosine is present in a non-template strand (Beletskii and Bhagwat, 1996). These observations are consistent with the notion that R-loops promotes TAM. R-loops mediated TAM have also been established as intermediates of both DSBs and transcription-associated recombination (TAR) (Chavez and Aguilera, 1997, Huertas and Aguilera, 2003, Wimberly et al., 2013). Meanwhile, nucleotide excision repair (NER) proteins have been found to engage in the metabolism of R-loops in human cells into potentially dangerous DSBs (Sollier et al., 2014).

R-loops can also be deleterious by interfering with transcription. As described in Section 1.12, R-loops can participate in chromatin re-modelling that possibly interferes with global gene expression (Ginno et al., 2012, Castellano-Pozo et al., 2013, Skourti-Stathaki et al., 2014). This could affect the expression of repair and replication genes that could have far-ranging consequences for genome integrity. R-loops can also affect transcription directly by impeding its progression (Hamperl et al., 2017), affecting gene dosage. R-loop-dependent transcription silencing may also be important for the pathophysiology of diseases with repeat instability. Repeat instability refers to the characteristic of some DNA repeats to expand or be reduced, causing an unstable number of repeats. Usually, the instability involves expansion of trinucleotide repeats, perhaps as such mutations would not cause frameshifts. However, the repeats involved need not be limited to protein-coding regions (Li et al., 2008). Repeat instability leads to several diseases, including the Fragile X syndrome and Friedreich Ataxia (Castel et al., 2010). Interestingly, transcription of repeat nucleotides responsible for Friedreich Ataxia in *E. coli* led to R-loop formation that impeded further transcription (Grabczyk et al., 2007). In the same study, *in vitro* transcription of the same repeat nucleotides using the T7 RNA polymerase also led to R-loop formation and subsequent inhibition of transcription. More recently, R-loops were found to accumulate at repeat nucleotides in patients of both Friedreich Ataxia and Fragile X Syndrome, promoting gene silencing (Groh et al., 2014, Groh and Gromak, 2014). Whether R-loops actually cause the repeat instability is unknown but these findings suggest that it is involved in the pathophysiology of at least a sub-set of the associated disorders.

R-loops also contribute to genome instability by interfering with fork progression. Although the replication process itself seem to be important for generating or stabilizing R-loops on the genome (Hamperl et al., 2017), R-loops may strengthen the interaction between the transcription machinery and its DNA substrate, impeding fork progression. Whether these collisions involve direct physical contact between the replication and transcription machineries or are mediated indirectly through super-coiling in the DNA substrate is subject to debate. Another plausible scenario is that TAM, TAR and DSBs generated

by R-loops (or the repair proteins recruited as a consequence) could act as fork barriers. In fact, R-loops are enriched on longer genes in human cells at common fragile sites (CFSs), loci characterized as hotspots for DNA breaks and where collisions between replication forks and transcription complexes occur at anomalously high frequencies (Helmrich et al., 2011). Consequently, R-loops could contribute to the pathophysiology of diverse ailments, including Duchenne and Becker muscular dystrophy as well as juvenile Parkinsonism as the genes responsible for these diseases (DMD and PARK2 respectively) are both characterized by high numbers of CFSs (Mitsui et al., 2010). Alternatively, R-loops could impede fork progression independently of the transcription machinery or other protein or protein-nucleic acid complexes. Whatever the precise context through which R-loops inhibit fork progression, fork stalling and collapse might ensue, leading to genomic instability. Finally, persistent R-loops at the *rDNA* locus in yeast can be used as primers for the initiation of non-canonical DNA replication, contributing to dangerous re-replication events (Stuckey et al., 2015). It should be stressed, however, that R-loops by and of themselves are not sufficient for genome instability (García-Pichardo et al., 2017). This was demonstrated when the authors identified histone mutants that led to accumulation of R-loops, but with limited genome instability, unlike in *hpr1Δ* and *sen1-1* cells. Moreover, these mutants led to reduced levels of H3S10 phosphorylation and also suppressed the levels of genomic instability in both *hpr1Δ* and *sen1-1* cells (García-Pichardo et al., 2017). The authors concluded that, upon formation of R-loops, a second epigenetic step (either limited to or including H3S10 phosphorylation) was necessary for R-loop mediated genome instability.

Given the various ways through which R-loops can compromise genome integrity, it is not surprising to note that cells have evolved several pathways to rid the genome of R-loops when required. In fact, cells seem to adopt a simplistic approach to dealing with R-loops. Firstly, mechanisms are in place to prevent the enrichment of R-loops beyond levels that are physiologically beneficial for the cells. On the other hand, once stable R-loops have been generated, mechanisms distinct form those involved in R-loop prevention are utilized to remove them from the genome (Fig 1.12).

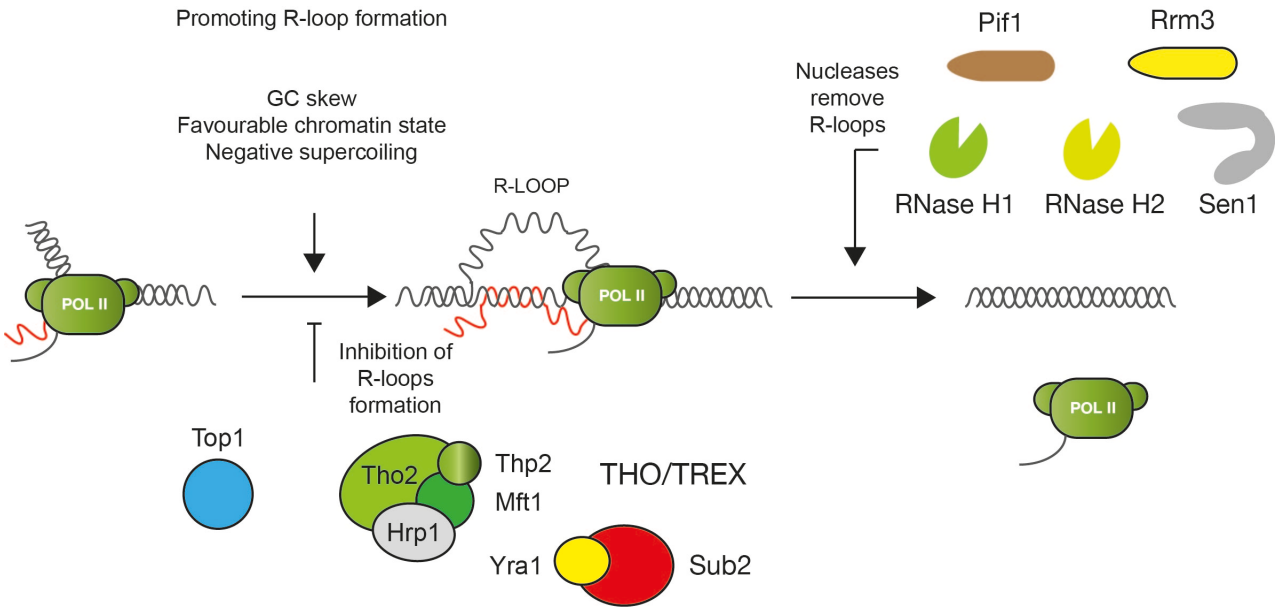


Figure 1.12. Summary of factors that are important for the homeostasis of R-loops. The nature of the DNA being transcribed is important for R-loop formation. For instance, GC skew favours the formation of R-loops. Moreover, DNA left behind the transcription machinery is negatively supercoiled, also favouring R-loop formation. Because of the potentially deleterious consequences of stable R-loops in the genome, cells employ several pathways to prevent R-loop formation in the first place. For instance, topoisomerases alleviate negative supercoiling behind RNA polymerases. Nascent transcripts are also efficiently exported out of the nucleus using the THO/TREX complex, preventing R-loop formation. Once R-loops are formed, however, cells have at their disposal several nucleases that they employ to remove them. These include the RNase H endonucleases and the conserved DNA/RNA helicase, Sen1.

Efficient processing of transcripts is a major factor in preventing the formation of R-loops. In *S. cerevisiae*, deletion of the *HPR1* gene (part of the THO/TREX complex that is used to process transcripts out of the nucleus) led to the enrichment of R-loops *in vivo*. These cells were also characterized by hyper-recombination (Huertas and Aguilera, 2003). These observations highlight a possible link between stable R-loop formation and genome instability. In DT40 chicken cells, depletion of ASF/SF2 (an enzyme involved in transcript splicing) by using a tetracycline-repressible promoter led to a subpopulation (~10%) of cells acquiring tetracycline resistance (Li and Manley, 2005). Further analysis revealed that tetracycline insensitivity was induced by R-loop dependent hyper-recombination that also caused gene expression from an inactivated promoter. Another line of evidence highlighting the role of efficient transcript processing in the prevention of R-loop formation comes from studies in the pathophysiology of some cancers. This includes cancer-formation in immuno-depressed individuals infected by the Kaposi's sarcoma-associated herpesvirus (KSHV). Studies suggest that a viral protein from KSHV impedes normal functioning of the human TREX complex, leading to stable formation of R-loops that contributes to DSBs (Jackson et al., 2014). Another example involves FIP1L1, an oncogene involved in leukaemia (Stirling et al., 2012). The FIP1L1 gene-product is a subunit of the human cleavage and polyadenylation specificity factory (CPSF) that is required for 3'-end processing and transcription termination. Depletion of FIP1L1 causes chromosome breakage, possibly due to R-loop enrichment (Stirling et al., 2012). Importantly, truncation mutants as well as temperature-sensitive mutants (grown at non-permissive temperatures) of the yeast homolog of *FIP1L1* (*FIP1*) lead to both enrichment of R-loops and increased genomic instability (Stirling et al., 2012).

Negative super-coiling favours R-loop formation (Roy et al., 2010). *In vivo*, this favourable coiling arises behind transcription complexes. Unsurprisingly then, topoisomerases have been implicated in the prevention of R-loop formation. Indeed, in yeast, absence of topoisomerases Top1 and Top2 led to R-loop enrichment at the *rDNA* locus (El Hage et al., 2010). Moreover, R-loop enrichment at that locus in *top1*Δ *top2*Δ double mutant cells interfered with RNA Pol I transcription. In human cells, treatment with topoisomerase

inhibitors (campothecin and topotecan) results in accumulation of R-loops at the appropriate DNA substrates (Powell et al., 2013, Sollier et al., 2014, Sordet et al., 2009).

Once R-loops have been formed, they need to be removed from genomes. One of the main pathway of R-loop removal involves the two RNase H endonucleases. The substrate requirements between RNase H1 and H2 differ. Current evidence suggests that RNase H1 recognizes and binds to both the RNA and DNA portion of DNA/RNA hybrids but it requires the hybrids to comprise of at least four consecutive ribonucleotides in order to process them. On the other hand, RNase H2 can recognize and cleave single ribonucleotides incorporated into the genome (Cerritelli and Crouch, 2009). Thus, both enzymes can remove R-loops. In fact, in mutant cells that otherwise stabilize R-loops, overexpression of RNase H1 or H2 can suppress R-loop accumulation (Huertas and Aguilera, 2003, Mischo et al., 2011). Moreover, sensitivity to RNase H treatment *in vitro* is a diagnostic for the presence of R-loops (Li and Manley, 2005, Roy et al., 2008, Roy and Lieber, 2009). In humans, mutations in the subunits of the RNase H2 enzyme lead to an autoimmune inflammatory disorder named Aicardi-Goutières syndrome (AGS). As such, a decrease in RNase H-mediated R-loop processing could contribute in part to the pathophysiology of AGS (Groh and Gromak, 2014). In fact, the autoimmune nature of AGS is linked to the inappropriate accumulation of nucleic acids that is thought to trigger aberrant type I interferon signalling and, hence, chronic inflammation (Crow et al., 2014).

R-loops are also targeted by specific helicases. In human cells, depletion of the DNA/RNA helicase Aquarius (AQR) led to the accumulation of R-loops and DSBs, indicating that Aquarius participates in R-loop removal (Sollier et al., 2014). Overexpression of RNase H1 reduces the enrichment of R-loops in cells depleted for the Aquarius. Interestingly, Aquarius is part of a subfamily of DNA/RNA helicases that possesses a DEAxQ-like catalytic domain. Another protein with this DEAxQ-like domain in humans is Senataxin (SETX) (Sollier et al., 2014). Depletion of Senataxin also causes R-loop enrichment that can be remediated by the overabundance of RNase H1 (Skourti-Stathaki et al.,

2011). Mutations in the Senataxin gene can lead to one of two neurological diseases; ataxia with oculomotor apraxia type 2 (AOA2) and juvenile-onset amyotrophic lateral sclerosis type 4 (ALS4). Whether R-loops contribute to these pathologies remain to be seen. Unlike AGS, neither AOA2 nor ALS4 have been recognized as autoimmune diseases yet. However, depletion of Senataxin in human cells have been shown to lead to increased transcription of antiviral genes upon infection of the cells with the influenza A virus (Miller et al., 2015). Cells derived from individuals with AOA2 and cells from *Setx*$^{-/-}$ mice were also characterized by increased transcription of anti-viral genes upon infection using the influenza A and the Sendai virus (murine para-influenza virus) defective-interfering RNA respectively (Miller et al., 2015). This may lead to chronic inflammation. In fact, both AOA2 patients and *Setx*$^{-/-}$ mice are hyper-responsive to infection, suggesting that chronic inflammation may contribute to the pathophysiology of AOA2 (Becherel et al., 2015). Meanwhile, a murine model of ALS (but not specific to ALS4) were characterized by the upregulation of interferon-stimulated genes in the astrocytes surrounding motor neurons (Wang et al., 2011), suggesting that chronic inflammation may also contribute to the pathophysiology of ALS4.

Senataxin, but not Aquarius, is conserved in eukaryotes. Its yeast orthologue, Sen1 (Splicing endonuclease 1) was originally isolated as a protein important in the homeostasis of tRNA precursors (DeMarini et al., 1992). Increasingly however, Sen1 is being regarded as an important player in both R-loop removal (Mischo et al., 2011) and in non-canonical transcription termination. Sen1 also functions at forks, at the interface of transcription and DNA replication (Alzu et al., 2012). In fact, the role of Sen1 at forks will be a major focus of this thesis. The different functions of Sen1 will be discussed in more detail in later sections. Presently, however, the mechanism of canonical transcription termination in eukaryotes will be described and this mechanism will later be contrasted with the non-canonical, Sen1-dependent mechanism of transcription termination.

1.14 CANONICAL TRANSCRIPTION TERMINATION IN EUKARYOTES

Transcription is the process through which RNA polymerases synthesize transcripts, using the genome as a template. Eukaryotic transcription differs tremendously from its prokaryotic counterpart in that it occurs in the nucleus, in the absence of the translation machinery. In eukaryotes, three RNA polymerases operate. These enzymes are related structurally and functionally but they have undergone specialization and each transcribes different RNA molecules using distinct loci of the genome as templates (Table 1.2)

Table 1.2. The eukaryotic RNA polymerases. Here, some of the most common products of the eukaryotic RNA polymerases are given (Richard and Manley, 2009). This list is non-exhaustive and the sedimentation of the corresponding rRNA molecules are different between yeast and mammals.

Enzyme	Transcripts
RNA Pol I	• 25-28S rRNA • 16-18S rRNA • 5.8S rRNA
RNA Pol II	• Spliceosomal small nuclear RNA (involved in splicing/RNA processing) • Small nucleolar RNA (a subtype of snRNAs involved in the chemical modification of RNA molecules, including rRNA, tRNA, and snRNA) • MicroRNA precursors (miRNA are involved in RNA silencing and post-transcriptional regulation of gene expression. • CUT (cryptic unstable transcripts, products of non-productive transcription). • Long non-coding RNA (regulation of gene expression) • Messenger RNA
RNA Pol III	• Transfer RNA • 5S rRNA • U6 spliceosomal snRNA

Independently of the identity of the transcripts being synthesized, transcription need to be terminated reproducibly. This is especially true for protein coding genes. Failure to do so will lead to read-through transcription. Beyond the obvious futile and costly energy expenditure in transcription beyond the limits of a certain gene of interest, read-through transcription from an upstream gene can proceed to the promoter of an immediately downstream gene, interfering with the activity of the second promoter. This phenomenon, called transcription interference, may result in decreased transcript levels from the downstream gene and correspondingly impact on the protein levels encoded by that gene, as evidenced by experiments in *S. cerevisiae* at the *GAL10* and *GAL7* genes (Greger and Proudfoot, 1998). Notably, the two genes are transcribed co-directionally and the *GAL7* gene is found downstream of the *GAL10* gene. Deletion of the termination sequence of the *GAL10* gene led to complete shut-off of *GAL7* expression whilst, concomitantly, a bi-cistronic *GAL10-GAL7* transcript was highly enriched.

Transcription from a certain gene will often generate several mRNA populations varying in length. Some of these isoforms arise from alternative splicing whilst others are a product of using different polyadenylation signals (PASs) that signal for transcription to be terminated (Proudfoot, 2011). Indeed, mutations that inactivate PASs lead to transcription hundreds of bases beyond the PAS, leading to suboptimal levels of the genes being transcribed. In fact, this defect underpins the pathophysiology of two different thalassaemias (blood disorders characterized by incorrect blood counts) (Higgs et al., 1983, Orkin et al., 1985, Whitelaw and Proudfoot, 1986).

Passage through PASs is sensed by RNA Pol II. In fact, RNA Pol II differs from either RNA Pol I and III by virtue of an additional protein segment that is formed from the carboxyl-terminal domain (CTD) of its largest subunit, Rpb1 (Vannini and Cramer, 2012). This CTD consists of repeats of a seven residue-stretch of amino acid (consensus Tyr^1-Ser^2-Pro^3-Thr^4-Ser^5-Pro^6-Ser^7) and the number of these repeats varies from 26 in *S. cerevisiae* and 29 in *Sz. pombe* to 52 in vertebrates (Schwer and Shuman, 2011). Using recombinant expression of murine Pol II in human cells, it was found that truncating the

number of the heptad repeats from 52 down to five resulted in decreased transcription termination (McCracken et al., 1997). The authors also found that the heptads were required for binding to components of the cleavage and polyadenylation (CPA) complex. Importantly, the different residues within the heptad can be modified and different modifications are thought to enable distinct interactions with diverse complexes and to mediate distinct molecular events (Buratowski, 2003). Canonical transcription termination of Pol II genes *in vivo* seems to depend on concomitant de-phosphorylation of Tyr^1 (Schreieck et al., 2014) when RNA Pol II reaches a PAS and continued Ser^2 phosphorylation (Ahn et al., 2004). These modifications seemingly enable recruitment of Pcf11 (a component of the CPA complex) and Rtt103 (Kub5-Hera in humans) that participate in 3'-end processing and transcription termination respectively (Schreieck et al., 2014). It is likely however that the effects of modifications of the heptad repeats on transcription termination are nuanced. For instance, in *Sz. pombe*, mutagenesis of Tyr^1 to either alanine or leucine is lethal but mutation to its non-phosphorylable surrogate, phenylalanine, is not (Schwer and Shuman, 2011). This suggests that Tyr^1 phosphorylation is not an absolute requirement for termination. Likewise, mutagenesis of Ser^2 to alanine or threonine was viable but mutation to the phospho-mimetic glutamic acid was found to be lethal. Although the viable mutants were temperature-sensitive, suggesting some defects, these results indicate that Ser^2 phosphorylation is not necessary for the termination of transcription of RNA Pol II genes.

Transient pausing of RNA Pol II after passing the PAS has also been implicated in transcription termination and this pausing has been shown to be dependent on the PAS itself and not on other *cis*-acting elements (Orozco et al., 2002). This pausing is stochastic and leads to the gradual termination of transcription. Pausing may also be dependent on the recruitment of the CPA complex at or near the PASs. Indeed, depletion of components of the CPA complex in human cells leads to defects in pausing at PASs (Nojima et al., 2015). R-loops and heterochromatin may also contribute to pausing of the RNA Pol II enzyme (Proudfoot, 2016). Stalling of RNA Pol II is thought to coincide with conformational changes in the protein, prompting transcription

termination. The CPA then polyadenylates the 3'-end of the transcript, leading to mature mRNA. These molecules can then be exported out of the nucleus whilst the RNA Pol II can be recycled and used anew for transcription. This model has been dubbed the **allosteric** model of termination (Fig 1.13).

An alternative pathway for canonical transcription termination involves the Rat1 5'-3' endonuclease in budding yeast (Dhp1 in fission yeast and XRN2 in vertebrates). In this model, the 3'-end of the nascent transcript is still cleaved when RNA Pol II passes through a PAS and the transcript is polyadenylated. However, in this model, RNA Pol II goes on to transcribe beyond the PAS. Simultaneously, the Rat1 endonuclease is recruited at the PAS and degrades the read-through transcript until it catches up with the processive RNA polymerase and **torpedoes** it off the chromosome (Fig 1.13). Indeed, yeast strains carrying the temperature-sensitive *rat1-1* allele grown at non-permissive temperatures are characterized by transcription termination defects and increased stability of transcripts downstream of their PASs (Kim et al., 2004). Likewise, *rai1*Δ cells (Rai1 forms a complex with Rat1) are also characterized by termination defects and enhanced stability of transcripts beyond their PAS (Kim et al., 2004). In HeLa cells, depletion of XRN2 using the siRNA technology also leads to termination defects (West et al., 2004). These observations suggest a non-species specific function for Rat1/XRN2 in transcription termination.

Back-tracking may play an intriguing role in transcription termination. RNA Pol II has an intrinsic exonuclease activity to which the transcription factor TFIIS is a co-enzyme (Proudfoot, 2016). Backtracking favours removal of incorrectly transcribed RNA molecules but the enzyme might encounter secondary structures within the transcript during back-tracking that promotes transcription termination. *In vitro* experiments using endogenous *S. cerevisiae* RNA Pol II and recombinant Rat1-Rai1 indicate that misincorporation of dNTPs in the transcript by RNA Pol II promotes Rat1-Rai1 dependent degradation, suggesting a role for backtracking in termination (Park et al., 2015). Interestingly, substituting Rat1-Rai1 with human XRN2 also leads to degradation of nascent transcripts upon misincorporation of dATP. Given the

similarity between RNA Pol I, II and III, it is interesting to note that Rat1 is required for productive termination of transcription of RNA Pol I although the latter lacks a CTD (El Hage et al., 2008) whilst backtracking has been implicated in termination of transcription of RNA Pol III (Nielsen et al., 2013).

It is possible that the allosteric and alternative (torpedo) models of transcription termination are mediators of the same (canonical) pathway. In this scenario, choosing one pathway of termination over the other would depend on some threshold being crossed or reached. For example, it is possible that, when RNA Pol II does not pause sufficiently at several PASs, the Rat1/Rai1 complex is recruited instead. As such, at any particular locus, either model of canonical termination could be used. It should be noted however that transcription can be terminated in a non-canonical manner. Often, this will involve abrupt arrest of RNA Pol II instead of pausing. Such arrests have been triggered by chemical damage. Once the RNA Pol II has been arrested, the enzyme is poly-ubiquitylated and targeted for degradation by the 26S proteasome (Svejstrup, 2007). The arrest also acts as a signal for transcription-coupled repair (TCR) using the nucleotide-excision repair (NER) pathway (Svejstrup, 2002). Although it is not known whether such arrests do occur under physiological conditions, it has been proposed that, in the event of collisions between forks and R-loop stabilized transcription complexes, RNA Pol II arrest could be used to forcefully terminate transcription (Proudfoot, 2016). Interestingly, in yeast, a protein involved in non-canonical transcription termination, has been shown to associate with forks (Alzu et al., 2012). The protein in question is the DNA/RNA helicase, Sen1 (Fig 1.13).

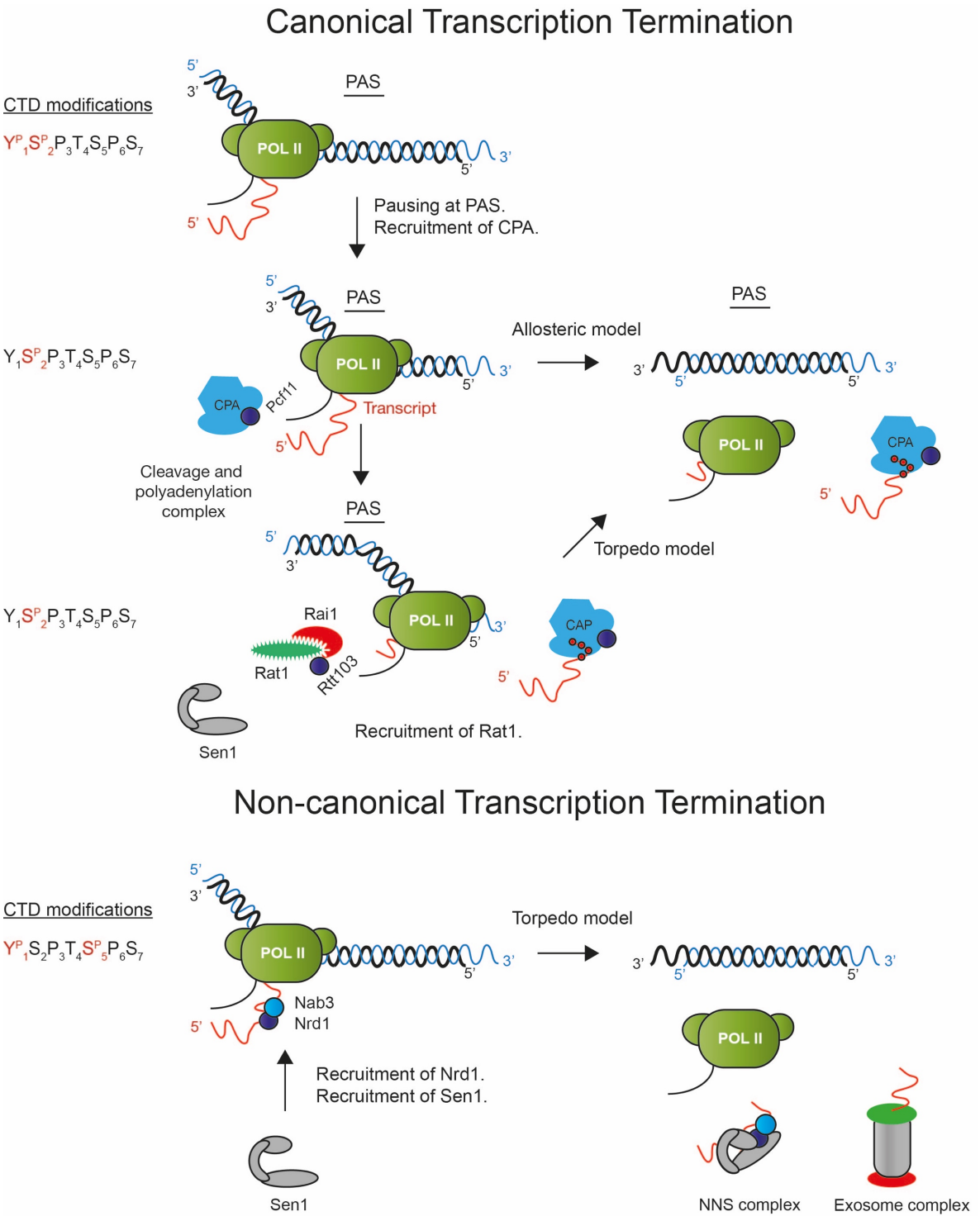


Figure 1.13. In *S. cerevisiae*, cells can terminate RNA Pol II-mediated transcription in one of two ways. The main mechanism involves the cleavage and polyadenylation of the nascent transcript once the RNA polymerase has reached a terminator sequence named PAS. Passage through the PAS coincides with changes in the conformation of the enzyme that can no longer bind to its substrate. The enzyme is then recycled. If the polymerase does not disengage, the Rat1 endonuclease is recruited and the latter torpedoes the polymerase off the genome. An alternative, non-canonical

pathway to terminate transcription involves the NNS complex where Nrd1 and Nab3 enable recruitment of the DNA/RNA helicase that torpedoes the polymerase, similar to Rat1 (Porrua and Libri, 2013). The NNS interacts with the exosome and this interaction may be important in NNS-dependent transcription termination. Interestingly, Sen1 may also be involved in canonical transcription, independently of Nrd1 and Nab3.

1.15 NON-CANONICAL/ PREMATURE TRANSCRIPTION TERMINATION BY THE NNS COMPLEX IN *S. CEREVISIAE*

In *S. cerevisiae*, transcripts synthesized by RNA Pol II shorter than 1000 bases are typically terminated by a mechanism that is not dependent on the cleavage and polyadenylation of the 3'-end of the transcript. Instead, these transcripts are terminated by the NNS complex made up of Nrd1, Nab3 and Sen1. The substrates of the NNS termination pathway include snRNA (Steinmetz et al., 2001), snoRNA (Kim et al., 2006) and cryptic unstable transcripts (CUTs) (Arigo et al., 2006), although it is likely that several substrates of the NNS complex remain to be identified as such.

Nrd1 and Nab3 are two sequence-specific RNA binding proteins that form a heterodimer on the nascent transcript (Nrd1 binds to **GUAA** and Nab3 binds to **UCUU**) (Carroll et al., 2007, Conrad et al., 2000). Single-base mutations of the Nrd1 and Nab3 binding sites to **G*C*AA** and **UC*G*U** respectively were sufficient to reduce the binding affinity of the Nrd1-Nab3 dimer to its RNA template 10 to 20-fold *in vitro* (Carroll et al., 2007). *In vivo*, these mutations lead to termination read-through. As such, the ability of the Nrd1-Nab3 dimer to dock onto nascent transcripts is critical for the NNS-dependent transcription termination. The primary sequence of the N-terminal region of Nrd1 has similarity with the RNA Pol II CTD-interacting domains (CIDs) of both Pcf11 and Rtt103 and was shown to also interact with the CTD of RNA Pol II (Vasiljeva et al., 2008) and the *nrd1-101* allele, that encodes for a mutation in the CID of Nrd1, leads to termination defects (Conrad et al., 2000). Unlike Pcf11 and Rtt103, however, Nrd1 preferentially interacts with the RNA Pol II CTD when the Ser^5 but not the Ser^2 is phosphorylated (Vasiljeva et al., 2008). As such, occupancy on RNA Pol II by Nrd1 and by either Pcf11 or Rtt103 would seem to be mutually exclusive. Indeed, a simplistic model suggests that Ser^5 is phosphorylated when RNA Pol II is close to the 5'-end of genes but, as it elongates, Ser^5 is dephosphorylated whilst Ser^2 is phosphorylated (Hsin and Manley, 2012). To complicate matters however, the Nrd1-Nab3 dimer can bind to the 3'-UTR (untranslated regions) of several genes (Webb et al., 2014).

Importantly, neither Nrd1 nor Nab3 can bind stably to mRNA in isolation of one another (Carroll et al., 2007) and removal of the residues of Nrd1 responsible for interaction with Nab3 abrogates interaction between the CTD of RNA Pol II and Nrd1 (Vasiljeva et al., 2008). Intriguingly, neither the deletion of Nrd1's CID nor that of the residues required for interaction with Nab3 cause lethality in *S. cerevisiae*. However, the *NRD1* gene is essential for viability and deletion of both domains without interfering with the RNA-binding domain is lethal suggesting that these two domains serve to perform overlapping, if not entirely redundant, functions (Vasiljeva et al., 2008).

Nab3 is also essential for viability but its function remains unclear. Whether Nab3 is functional in isolation is also unclear. Work in the Corden lab has shown that Nab3 overexpression correlates with increased phosphorylation of Nrd1 possibly as Nrd1 needs to form a dimer in order to be phosphorylated (Conrad et al., 2000). Yeast-two-hybrid experiments have also shown that Nab3, but not Nrd1, can physically interact with a Sen1 construct spanning residues 1890 to 2092 (Nedea et al., 2008). As such, current evidence points to a direct interaction between Nab3 and Sen1, forgoing any such interaction with Nrd1. This interaction is important for recruitment of the Sen1 protein at nascent transcripts. Moreover, immuno-precipitations suggest that full-length Sen1 forms a complex with Nrd1, Nab3 and Glc7 (a subunit of CPF, the cleavage and polyadenylation factor, used in mRNA cleavage and polyadenylation) (Nedea et al., 2008). Nrd1-Nab3 forms a stable sub-complex within this tetrameric complex and may serve to stabilize it.

The final component of the NNS complex is Sen1. The *SEN1* gene encodes a 253 kDa (2231 amino acid) DNA/RNA helicase that belongs to the Upf1-like superfamily 1 helicases (Jankowsky, 2011). Deletion of the gene is lethal in yeast and the essential region has been mapped to the C-terminal domain of the protein (DeMarini et al., 1992). This region encodes the helicase domain of the protein. The precise mechanism through which Sen1 mediates termination is still being elaborated. An elegant *in vitro* study has demonstrated that Sen1 disengages the transcription machinery in a manner reminiscent of the bacterial Rho termination factor (Porrua and Libri, 2013).

Indeed, Sen1 binds to transcripts and uses its ATPase activity to translocate along the RNA molecule in the 5' to 3' direction until it reaches the DNA/RNA region. There, it unwinds the transcript from its template. This termination was not dependent on the presence of Nrd1 and Nab3 and was carried to completion whether the RNA Pol II was actively transcribing or was stalled by the Reb1 protein. Using non-hydrolysable ATP or a mutant of Sen1 incapable of ATP-hydrolysis (Sen1G1747D) did not lead to productive transcription termination suggesting that ATP hydrolysis was important for Sen1-dependent transcription termination. Meanwhile, both recombinant full-length Sen1 and Sen1 (1095-1876) have demonstrable helicase activity *in vitro* and are capable of unwinding both RNA and DNA substrates (Martin-Tumasz and Brow, 2015, Leonaitė et al., 2017). As such, Sen1 may mediate two possibly-distinct enzymatic activities from its helicase domain. At non-permissive temperatures, *sen1-1* strains exhibit strong termination defects and increased levels of R-loops. These defects were alleviated by expression of Sen1 constructs expressing the helicase domain expressed under the control of the strong *ACT1* promoter (Mischo et al., 2011). Which of the helicase or transcription terminator functions of Sen1 is absolutely necessary for cell survival and whether these functions can be separated genetically are important avenues of future research.

To enable transcription termination, Sen1 needs to be brought close enough to a transcribing RNA Pol II enzyme. However, the mechanism through which Sen1 interacts with RNA polymerases is not clear. It has been established that, like the Rat1 endonuclease, Sen1 is able to terminate transcription mediated by RNA Pol II but not by *E. coli* RNA polymerase (Park et al., 2015, Porrua and Libri, 2013). This indicates species- or polymerase-specificity. Surprisingly, however, using a variant of RNA Pol II lacking its CTD did not impair Sen1-dependent termination *in vitro* (Porrua and Libri, 2013). This is at odds with yeast-two-hybrid assays that indicate that Sen1 interacts preferentially with phosphorylated Ser^2 in the CTD of RNA Pol II, independently of Nrd1 and Nab3 (Chinchilla et al., 2012). Indeed, a Sen1 mutant (Sen1R302W) that has decreased affinity to phosphorylated Ser^2 in the CTD of RNA Pol II has a different Chip-chip trace compared to wildtype

Sen1 at both non-coding and protein-coding loci. Taken together, this suggests that Sen1 interacts with the CTD terminal of RNA Pol II but that it can also interact with a non-CTD domain of RNA Pol II. In fact, IPs experiment have shown that Sen1 interacts with RNA Pol I, II and III (Yüce and West, 2013). Unlike RNA Pol II, both RNA Pol I and III lack a CTD. This strengthens the hypothesis that Sen1 interacts with RNA Pol II in at least two different ways.

Transcription speeds play a critical part in Sen1-dependent transcription termination, as illustrated in (Hazelbaker et al., 2013). The authors used two mutant alleles of *RPB1* (the gene that encodes the largest subunit of RNA Pol II): *rpb1E1103G*^{FAST} that leads to increased transcription speeds and is characterized by increased defects in transcription termination and *rpb1N488D*^{SLOW} that is characterized by reduced transcription speeds with no defects in transcription termination (Hazelbaker et al., 2013). Temperature sensitive-alleles of *SEN1* (*sen1-1* and *sen1-E1597K*) were shown to have different responses to these two mutant alleles of *RPB1*. At semi-permissible temperatures, *rpb1N488D*^{SLOW} was shown to suppress the growth defects associated with *sen1-1* and *sen1-E1597K* whilst *rpb1E1103G*^{FAST} accentuated the growth defects of the *SEN1* mutants. Moreover, at semi-permissive temperatures, *rpb1N488D*^{SLOW} suppressed the termination defects of the *SEN1* mutants whilst *rpb1E1103G*^{FAST} led to increased termination defects. The increased defects can be accounted for by additive effects of the *rpb1E1103G*^{FAST} and *SEN1* mutants. However, the synthetic suppression of *SEN1* mutants by *rpb1N488D*^{SLOW} suggests that there is an interplay between the speeds of RNA Pol II and Sen1-dependent transcription termination.

Several interactors of Sen1 could play a role in NNS-dependent transcription termination. Sen1 interacts physically and genetically with Rnt1, a homolog of *E. coli* RNase III, that is required for processing the 3'-end of U2 and U5 spliceosomal RNA (Chanfreau et al., 1997, Elela and Ares, 1998, Ursic et al., 2004). In fact, Rnt1 promotes transcription termination by enabling Rat1 to access RNA Pol I (El Hage et al., 2008, Kawauchi et al., 2008) and Pol II (Ghazal et al., 2009). As such, Sen1 could compete with Rat1 for Rnt1-

dependent recruitment to RNA polymerases. Both Rat1 and Sen1 have been found to be important for transcription termination of RNA Pol II at the PAS of a reporter gene. Moreover, cells carrying both the *rat1-1* and *sen1-1* alleles showed additive defects in transcription termination at PAS (Kawauchi et al., 2008). This suggests that Sen1 is able to terminate transcription both though the canonical and non-canonical pathways.

Sen1 also interacts with Glc7 (Nedea et al., 2008). Sen1 is thought to present Glc7 to the Nrd1-Nab3 dimer, in proximity to nascent transcripts. In yeast, the CPF consists of several sub-complexes, one of which is known as the APT (Associated with Pta1) complex (Nedea et al., 2003). APT is made up of seven proteins, namely: Glc7, Pta1, Pti1, Ref2, Ssu72, Swd2 and Syc1. Deletion of either *SWD2* and *REF2* destabilizes the APT complex so that Glc7 no longer associates with the complex, triggering cell lethality. Over-expression of Sen1 can synthetically suppress the lethality associated with *swd2*Δ and *ref2*Δ in W303 strains (Nedea et al., 2008). Importantly, although overexpression of Sen1 constructs suppress lethality in *swd2*Δ and *ref2*Δ cells, they do not prevent the destabilization of the APT complex. Moreover, suppression of lethality was dependent on residue F256 found in the C-terminal region of Glc7 and a KHVCF motif in Sen1 spanning residues 1999-2003 (Nedea et al., 2008). These allow for a direct physical interaction between Sen1 and Glc7, enabling the formation of a stable tetramer made up of the NNS complex and Glc7. This may serve the purpose to bring Glc7 to the vicinity of transcribing polymerases. Glc7 is a phosphatase that could potentially dephosphorylate Sen1, contributing to Sen1-dependent termination. Elsewhere, Glc7 has been shown to dephosphorylate Tyr^1 in the CTD of RNA Pol II. Failure to dephosphorylate Tyr^1 prevents recruitment of termination factors such as Pcf11 and Rtt103, contributing to termination defects (Schreieck et al., 2014). Finally, Sen1 physically interacts with Smd3, a protein involves in intron splicing (Fromont-Racine et al., 1997). It is unknown whether this interaction is useful for transcription termination.

As mentioned earlier, *sen1-1* strains are characterized by the enrichment of R-loops. These trigger hyper-recombination at significantly higher levels than

in *hpr1*Δ and *rat1-1* cells at semi-permissive temperatures (Mischo et al., 2011). By using different direct-repeat recombination substrates, the extent of recombination was also found to be dependent on both the length of the substrate (or duration of transcription) and the rate of transcription (Mischo et al., 2011). As such, transcription termination defects in the presence of Sen1-1 is favourable for R-loop enrichment on the genome. This illustrates how disrupting the terminator function of Sen1 can lead to genomic instability.

1.16 THE EXOSOME MAY BE IMPORTANT FOR NNS-DEPENDENT TRANSCRIPTION TERMINATION

The NNS complex interacts with the exosome complex. The latter is conserved and exists in both the cytoplasm and nuclei of eukaryotes. It is important for RNA metabolism and homeostasis (Bonneau et al., 2009). The core cytoplasmic exosome complex is composed of 10 essential subunits, including the 3'-5' RNase Dis3 (Chlebowski et al., 2013) whilst the nuclear exosome complex is made up of those 10 subunits and another 3'-5' RNase, Rrp6 (Chlebowski et al., 2013). Besides the RNase subunits, the exosome complex is made up of a hexameric core (composed of Mtr3, Rrp41, Rrp42, Rrp43, Rrp45 and Rrp46) onto which sits a cap composed of Csl4, Rrp4 and Rrp40. ssRNA is treaded into the channel formed by the core, that presumably allows for optimal enzymatic activity from the RNase subunits (Chlebowski et al., 2013).

Interaction of the NNS with the exosome complex is mediated chiefly through Nrd1. Nrd1 co-precipitates with several components of the RNA exosome complex, including Rrp6 (Vasiljeva and Buratowski, 2006). Notably, the interaction between the exosome and Nrd1 depends on the latter's CID, suggesting a link between Nrd1's ability to bind to RNA Pol II and its ability to target transcripts for exosomal degradation/processing (Heo et al., 2013, Kim et al., 2016). Moreover, two cofactors of the exosome, Mpp6 and Trp4 interact with Nrd1 in a mutually exclusive fashion (Kim et al., 2016). Trp4 is part of the TRAMP complex (Trp4/5-Air1/2-Mtr4 polyadenylation) and stimulates the exosome into either trimming or degrading RNA in a Rrp6-dependent (Kilchert et al., 2016). Interestingly, overexpression of Nab3, but neither of Nrd1 nor of Sen1, supresses thermosensitive *air1/2* mutants at semi- and non-permissive temperatures, and Nab3 also interacts with Rrp6 independently of Nrd1 (Fasken et al., 2015). This suggests that the whole NNS complex and not Nrd1 alone is important to couple transcription termination to exosomal processing of transcripts.

It is likely that targeting transcripts for exosomal processing is an important step for transcription termination. Indeed, delayed processing of transcripts could favour on-going transcription. This was illustrated in yeast cells where *rrp6*Δ leads to defects in NNS-dependent transcription termination at several loci, including those transcribed by RNA Pol II (Fox et al., 2015).

Taken together, the NNS complex likely participates in transcription termination through several, perhaps overlapping pathways. The NNS complex is firstly a co-factor for the exosome in *S. cerevisiae*, whereby Nrd1 and Nab3 bind to transcripts and target them for degradation or an alternative form of processing. It is not known whether transcription needs to be stalled in order for the NNS to act as an exosomal co-factor but it would seem likely. If the transcripts are not degraded promptly, this may lead to the resumption of transcription.

1.17 CONSERVATION OF THE SUBUNITS OF THE NNS COMPLEX

The Sen1 protein is expressed at low levels in cells and Nrd1 and Nab3 docking onto transcripts would enable localized and fine-tuned recruitment of Sen1 onto transcripts. Overexpression of Nrd1 leads to increased premature termination, possibly by over-enriching for Sen1 (Arigo et al., 2006). Once recruited at actively-transcribing sites, Sen1 can torpedo RNA Pol II off the chromosome, by interacting with the CTD of RNA Pol II. Alternatively, Sen1 could make use of its DNA/RNA helicase activity to remove both R-loops and transcripts from their templates. The complex function of the NNS is reflected by the fact its individual components have different localization profiles on chromosomes (Alzu et al., 2012, Webb et al., 2014). Moreover, point mutants of Nrd1, Nab3 and Sen1 have different transcriptomes from their wildtype counterparts and amongst themselves (Chen et al., 2017). Whilst some of these differences could be ascribed to the severity and penetrance of the different mutations, their effect on protein stability and differing effects on complex formation, at least some of the differences reflect the non-overlapping functions of Nrd1, Nab3 and Sen1 in transcription termination (Chen et al., 2017).

It should be noted that the components of the NNS complex are conserved. In *Sz. pombe*, the Sen1 protein has two orthologues; Dbl8 (~222 kDa) and Sen1$^{Sz.pombe}$ (~193 kDa), both of which share 31% sequence identity with *S. cerevisiae* Sen1. *dbl8*Δ and *sen1*Δ are viable in fission yeast, perhaps as a result of redundancy. Like its yeast orthologue, Sen1$^{Sz.pombe}$ has been shown to possess both DNA and RNA helicase activities (Kim et al., 1999). Additionally, Sen1$^{Sz.pombe}$ has been shown to alleviate topological stress in the genome at RNA Pol III genes in fission yeast by antagonizing Pol III-dependent transcription (Legros et al., 2014). Nrd1 and Nab3 are also conserved in fission yeast (Seb1 and Nab3, respectively) and both seem to have RNA recognition motifs (RRM), indicating a possible functional conservation. However, the RNA-binding protein Mmi1 seems to fulfil a role

comparable to that of the NNS complex in targeting mRNA molecules to the exosome complex in *Sz. pombe*, suggesting some divergence in functionality. In humans, the homolog of Sen1, Senataxin (SETX) (Chen et al., 2004) binds to the RRP45/EXOSC9 subunit of the human exosome complex in a SUMOylation-dependent manner. Mutations that impede the SUMOylation of SETX leads to a recessive neurological disorder, known as AOA2 (Richard et al., 2013). Mutations of SETX can also lead to a different neurological disorder known as ALS4. Both AOA2 and ALS4 will be discussed in more detail in Section 1.20. Paralogues of Senataxin include Aquarius, IGHMBP2, RENT1 and ZNFx1 (Bennett and La Spada, 2015, Chen et al., 2004, Sollier et al., 2014).

SCAF8/RBM16 is the likely mammalian homolog of Nrd1 and has been found to promiscuously interact with both phosohorylated Ser^2 and Ser^5 at the CTD of RNA Pol II (Becker et al., 2008, Corden, 2013, Patturajan et al., 1998, Yuryev et al., 1996). The human RALY protein's RRM share 31% identity with that of Nab3 and overexpression of the RALY gene, similarly to that Nab3 in yeast, can suppress thermosensitive *air1/2* mutants (Fasken et al., 2015). The biological function of the RALY protein has remained elusive however and polyadenylation-independent transcription termination as mediated by the yeast NNS complex has yet to be identified in humans (Corden, 2013). Like Sen1 however, SETX appears to be involved in canonical, Xrn2-dependent transcription termination (Skourti-Stathaki et al., 2011).

1.18 A PROMINENT ROLE FOR SEN1 IN TRANSCRIPTION-COUPLED REPAIR

Given its role in the removal of transcription-mediated R-loops and in transcription termination, it is not surprising to note that the Sen1 helicase is also involved in transcription-coupled repair. In fact, the first 975 residues of Sen1 was shown to interact with Rad2 in a yeast-two-hybrid assay (Ursic et al., 2004). Using the *sen1-1* and *rad2*Δ alleles, it was also shown that the two genes also interact genetically. Rad2 is a single-stranded DNA endonuclease involved in nucleotide excision repair (NER) (Habraken et al., 1993) where it is absolutely required for incision at the 3' end of lesions. NER, itself, is primarily important for the removal of UV-induced DNA damage such as pyrimidine dimers and can be sub-divided in two pathways: global genome NER (GG-NER) and transcription coupled repair (TCR). These differ in the substrate they recognize. GG-NER scans the genome for helix-distorting lesions whilst TCR primarily recognizes damage that impede RNA polymerase progression during transcription (Marteijn et al., 2014). However, after damage-recognition, the mechanisms of damage-removal and gap-filling in GG-NER and TCR are identical.

In *S. cerevisiae*, the DNA-dependent ATPase, Rad26, is important for TCR (van Gool et al., 1994). However, in the absence of TCR repressors such as Rpb4 (a dispensable subunit of RNA Pol II) (Li and Smerdon, 2002), as well as the transcription elongation factors Spt4 (Jansen et al., 2000) and Spt5 (Ding et al., 2010), Rad26 is dispensable for TCR. This suggests that there are at least two sub-pathways for TCR in budding yeast, one of which is dependent on Rad26. In fact, deletion of both *RBP9* (Rbp9 is another dispensable subunit of RNA Pol II) and *RAD26* abolishes TCR (Li and Smerdon, 2002), indicating that there are only two such sub-pathways. Interestingly, IPs of Sen1-TAP reveal that the protein interacts with elongation factors Spt5 and Spt6 and RNA Pol II subunits, including Rpb4 (Yüce and West, 2013).

Direct evidence for a role of Sen1 in TCR was provided when removal of either the first 1088 or last 373 residues of the protein led to defects in TCR (Li et al., 2016). Moreover, ablation of the last 373 residues led to milder TCR defects, suggesting that the N-terminal domain of Sen1 plays a more prominent role in TCR than its extreme C-terminal residues. The Sen1E1597K mutant, defective for its ATPase-helicase activity, only showed minor defects in TCR, suggesting that the catalytic activity of the protein is dispensable for its role in TCR. In addition, deletion of *SPT4* does not fully suppress the TCR defect associated with the *SEN1(1089-2231)* allele. This suggests that Sen1 works in the Rbp9-dependent pathway of TCR.

Sen1 could also play a role in non-TCR and non-NER repair pathways. Indeed, the triple *rad7*Δ *rad26*Δ *SEN1(1089- 1929)* mutant is more susceptible to UV compared to the double *rad7*Δ *rad26*Δ mutant. Rad7 is required for GG-NER. This suggests that Sen1 repairs UV-mediated damage through at least two repair pathways, one of which is not involved in NER. This role of Sen1 in repair of DNA damage may be conserved. Indeed, in humans, Senataxin co-localizes with 53BP1, a marker signalling response to DNA damage (Yüce and West, 2013).

1.19 SEN1 AT THE INTERFACE OF DNA REPLICATION AND TRANSCRIPTION

Besides its role in transcription, the human homolog of Sen1, Senataxin has also been implicated in DNA replication. In fact, both human epithelial U2OS cells and HeLa cells form nuclear Senataxin foci primarily when they replicate their genomes (Yüce and West, 2013). Moreover, treating cells with the B-DNA polymerase specific inhibitor aphidicolin, that works by docking at the active site of polymerases preventing dCTP incorporation (Baranovskiy et al., 2014), led to an increase in the number of Senataxin foci. Conversely, successive treatment of cells with aphidicolin and RNase H1 or treatment with the transcription-inhibitor α-amanitin led to a decrease in the number of Senataxin foci (Yüce and West, 2013). These observations indicate that Senataxin molecules are enriched on the genome especially when forks and transcription complexes are more likely to encounter one another and at sites of R-loop formation. Senataxin then may be important for fork progression past R-loops and transcription complexes.

In yeast, the first evidence for a possible role of Sen1 in DNA replication was provided when ChIP-chip analysis revealed that Sen1 is enriched at origins (Alzu et al., 2012). Moreover, at highly-transcribed RNA Pol II genes, such as *PDC1*, rapid depletion of Sen1 leads to enrichment of R-loops and Rfa1 (a subunit of the replication protein A) in the genome as determined by ChIP-qPCR. 2D gels also revealed the presence of gapped-forks following Sen1-depletion, indicative of incomplete DNA replication. As such, in yeast, Sen1 co-localizes with forks, enabling the completion of DNA replication. The authors also found that growing cells expressing the temperature-sensitive *sen1-1* allele at non-permissive temperatures led to the lengthening of S phase as well as terminal arrest in G_2. The *sen1-1* allele also triggers activation of Rad53 at non-permissive temperatures, suggesting that forks stall in the absence of a functional Sen1.

Sen1-1 also interacts genetically with several components of the replisome including *CTF4*, *MRC1* and *TOF1*, as well as genes involved in DNA repair including *RAD50* and *SGS1* (Alzu et al., 2012). Meanwhile, IPs of the CMG helicase has revealed that Sen1 interacts with the replisome (Giacomo De Piccoli, unpublished). Whether Sen1 directly binds to the replisome or interacts with the latter indirectly remains to be seen. The biological significance of this interaction is also an important aspect that needs addressing. These two questions will be a major focus of this thesis.

1.20 THE DOMAINS OF SEN1 AND SENATAXIN, AND THEIR ROLE IN HUMAN DISEASES

As mentioned earlier, the essential region of Sen1 has been mapped to the C-terminal domain of the protein (DeMarini et al., 1992) and this has been winnowed to residues 1084-1907, provided the presence of a chimeric nuclear localization sequence (NLS) (Chen et al., 2014). Analysis of the primary sequence of the C-terminal domain of Sen1 reveals that the portion of the protein spanning residues 1147-1857 has ~30% identical sequence with isoforms of its human homolog. This region corresponds to the helicase domain of the protein, suggesting some conservation of function. The crystal structure of this domain of Sen1 was recently published (Leonaitė et al., 2017). Like other Upf1-family helicases, Sen1 is made up of a subdomain 1B that forms a stalk and a barrel, and a subdomain 1C that forms a rigid prong. The latter extends atop a RecA1 domain. The authors suggest that once nucleic species enter between subdomains 1B and 1C, the subdomains close leading the prong to slice through the nucleic species, overcoming van der Waals forces between complementary bases, effectively causing melting. Interestingly, Sen1, contrary to other Upf1-like helicases, seem to have a unique structure that the authors call a brace that modulates the positioning of the barrel. Importantly, residues within this brace structure (1097-1149) are conserved in Senataxin (Leonaitė et al., 2017).

In silico analysis of the N-terminus of Sen1 suggests an abundance of α-helices. Whilst this domain is conserved in other yeasts (~31% identical sequence compared to both Sen1 and Dbl8 in *Sz. pombe*), it is not conserved in humans. However, this domain is conserved in aquatic vertebrates such as *Danio* and *Xenopus* (Bennett and La Spada, 2015). The N-terminal domain of Sen1 then underwent strong divergent evolution, especially in mammals. Given however that the N-terminal domain is not a separate protein, it is plausible that some of its functions are conserved (Bennett and La Spada, 2015). IPs of C-terminally FLAG-tagged Senataxin transiently transfected in HeLa cells revealed that the protein interacts with RNA Pol II subunits (RPB1,

RPB2 and RPB3), transcription elongation factors as well as with proteins involved in DNA repair such as MRE11 and RAD50 (Yüce and West, 2013). As such, like the N-terminal domain of Sen1, Senataxin can interact with RNA Pol II but, unlike Sen1, there is no evidence that Senataxin interacts with human RNA Pol I or III. Moreover, unlike Sen1, current evidence does not indicate Senataxin's involvement in TCR or NER in human cells.

Senataxin has been implicated in two distinct neurological disorders; ataxia with oculomotor apraxia type 2 (AOA2) and juvenile amyotrophic lateral sclerosis (ALS4) (Bennett and La Spada, 2015). AOA2 is an early-onset (10-22 years), progressive debilitating disease characterized by atrophied cerebellar matter, oculomotor apraxia (loss of or defective control of voluntary eye movements) and damage to the peripheral nervous system (Le Ber et al., 2004, Moreira et al., 2004). Of the original 15 mutations first described, 10 are nonsense mutations that lead to truncated proteins (Moreira et al., 2004), plausibly affecting the catalytic activity of the enzyme. Mutations that lead to AOA2 occur throughout the *SETX* gene, except at its extreme C-terminus (Chen et al., 2014). A study in the Brow lab exploited the conserved helicase domain of Sen1 to reproduce AOA2 point mutations in the yeast Sen1 protein (Chen et al., 2014). Of the 13 mutations tested, only two were indistinguishable from wildtype. The other mutations conferred either recessive or dominant transcription termination defects, temperature-sensitivity or were lethal. Four of those mutants were characterized by both temperature-sensitivity and defects in transcription termination. The N-terminal of Senataxin is SUMOlyated and this modification is a requisite for Senataxin's interaction with the exosome through its RRP45 subunit in human cells (Richard et al., 2013). This interaction is important for Senataxin to target the exosome to sites of DNA damage. Three AOA2 mutations were found to abolish the SUMOlyation of the N-terminus of Senataxin, thus inhibiting its interaction with RRP45. Notably, Senataxin and RRP45 colocalize at nuclear foci in S phase and these foci are sensitive to overexpression of RNase H1 (Richard et al., 2013, Yüce and West, 2013). Taken together, this suggests that AOA2 mutations are often loss of function mutations. Indeed, AOA2 mutations are recessive and carriers

do not present any phenotypes associated with the disease (Le Ber et al., 2004, Moreira et al., 2004).

Unlike AOA2, ALS4 is a dominant disease, suggesting gain of function mutations. ALS4 is much rarer than AOA2 potentially underlying a more specific and obscure pathophysiology. ALS4 is an early-onset disease, characterized by progressive degeneration of motor neurons in the cortex, brain stem and spinal cord. Analysis of unrelated pedigrees of the disease highlighted only four causal mutations: T3I, L389S and R2136C/H (Chen et al., 2004, Chen et al., 2014). Intriguingly, two of the mutations occur in the N-terminal domain of Senataxin whilst the other residue occurs in the helicase domain but is not conserved in Sen1 (Chen et al., 2014). The L389S mutation has been studied extensively. In a yeast-two-hybrid screen using a human brain cDNA library, and either Senataxin (1- 650) or Sentaxin^{L389S} (1- 650) as bait, the mutant fragment was seen to interact specifically with a peptide translated from an antisense transcript of a brain-specific, non-coding RNA known as BCYRN1 (Bennett et al., 2013). Whether this interaction is critical for the pathophysiology of ALS4 is yet to be determined. However, it is possible that ALS4 mutations lead to artificial interactors of the protein, directly antagonizing its useful, physiological roles in cells.

A *Setx* knock-out murine model was generated by crossing *Setx*+/- heterozygotes but this line of mice is both viable and displays neither ataxia nor neuronal degeneration phenotypes (Becherel et al., 2013). The male mice are infertile as a result of R-loops accumulating in germ cells, promoting apoptosis and hampering spermatogenesis. On the other hand, R-loops were not found to accumulate in post-mitotic cells (Yeo et al., 2014). This may underpin the discrepancy in the Senataxin-dependent development of neurological disorders between the murine model and humans. It is possible that, unlike in mice, R-loops could form in the human nerves encoding Senataxin mutants, contributing to the pathophysiology of AOA2 and/or of ALS4. Whilst not yet formally proven, it is possible that both AOA2 and ALS4 are characterized by chronic inflammation, similar to the autoimmune disorder, the Aicardi-Goutières syndrome.

1.21 AIMS OF THE THESIS

The goal of this thesis is to broaden our understanding of how forks deal with barriers in eukaryotes. For that purpose, I have used two different systems.

The first one involved the cell- or mating-type switching of *Sz. pombe*, where a fork barrier (*MPS1*) mediates fork pausing that is required for incorporating ribonucleotides at that specific locus (an event referred to as imprinting). Importantly, the Pol1 subunit of Polymerase α is required for this imprinting process. However, the mechanism through which Pol1 participates in the synthesis of the imprint has not yet been solved. Here, I have attempted to characterize the difference between *pol1* and its imprinting-defective mutant *swi7-1*. In order to do so, I have cloned recombinant variants of Pol1 and Swi7-1 and have assayed their properties *in vitro*.

The second system I was interested in involved the DNA/RNA helicase Sen1 travelling with forks in *S. cerevisiae*. Depletion of Sen1 leads to slow fork progression and activation of Rad53, suggesting that forks are stalled at a higher frequently in the absence of Sen1. Here, I sought to characterize the interaction between Sen1 and the replisome in *S. cerevisiae*. I have identified one replisome interactor of Sen1 as well as the domain through which Sen1 interacts with the rest of the replisome. I have also created an allele of *SEN1* that encodes for an isotype of the protein incapable of binding to the replisome in order to tease out the biological relevance of this interaction.

CHAPTER 2: MATERIALS AND METHODS

2.1 YEAST SPECIFIC METHODS

Both budding and fission yeasts are powerful model organisms for fundamental research. Of note, since the beginning of the 21st century, no less than six people have been awarded the Nobel Prize in Physiology or Medicine based on their work using either budding or fission yeast. They include Sir Paul Nurse and Leland Hartwell (2001) for their discoveries of key regulators of the cell cycle, Elizabeth Blackburn and Jack Szostack (2009) for their work on telomeres and telomerases, Randy Schekman (2013) for his involvement in the discovery of machinery regulating vesicle trafficking and Yoshinori Ohsumi (2016) for this work on autophagy.

Both species are attractive models for research for several reasons. Firstly, yeasts can be generated within a short time. Indeed, whilst human cell cycles last approximately 24 h, wildtype haploid strains of budding yeast complete an entire cell cycle within approximately 90 min in rich medium whilst wildtype haploid strains of fission yeast go through one cell cycle within 2-4 h. Consequently, a large number of cells can be generated relatively quickly for both organisms enabling rapid isolation of both nucleic acid and protein matter.

It is also easy and inexpensive to grow and maintain these two yeasts under laboratory conditions. Both organisms are widely used in research so that a wide range of techniques and reagents are available for use with either organism. Several techniques are applicable to both yeasts, with minor alterations required for optimal efficiency. Genetic manipulation is tractable, enabling gene deletion and disruption, as well as gene tagging with protein markers, degrons as well as reporter genes. Targeted mutagenesis is also possible. Unlike bacteria, budding and fission yeast can exist both as haploids and as diploids after mating between individuals of different and complementary mating/cell-types. As such, the phenotypes of recessive

alleles can be easily observed in haploid strains whilst heterozygous diploids deleted or disrupted for essential genes can also be constructed.

S. cerevisiae and *Sz. pombe* have both a rich history as model organisms in the study of DNA replication. Several proteins involved in the DNA replication process have been identified and characterized first in budding and fission yeast prior to the discovery of the human orthologue. Differences between the two yeasts and human cells do exist. For instance, neither budding nor fission yeast encode a geminin orthologue whilst *S. cerevisiae* has uniquely defined origins amongst eukaryotes. Notwithstanding such evolutionary quirks, the process of DNA replication is well conserved across eukaryotes so that studying the process in *S. cerevisiae* and *Sz. pombe* has direct implications for our understanding of DNA replication in humans. In this study, both yeasts were used as model organisms. Table 2.1 and 2.2 details the genotypes of the *S. cerevisiae* and *Sz. pombe* strains generated in this work respectively.

2.1.1 YEAST STRAINS AND MEDIA

S. cerevisiae strains were grown from 25% (v/v) glycerol suspensions (kept at −80°C) onto solid non-selective medium (YPD) at either 25°C for heat-sensitive strains or 30°C for all other strains. For *Sz. pombe* strains, cells were also grown from 25% (v/v) glycerol solutions (kept at −80°C) onto solid non-selective YEA medium at either 25°C for heat-sensitive strains or 33°C for all other strains. To select for antibiotics resistance conferred by the *kanMX* and *hphNT* cassettes, strains were grown on either YPD or YEA supplemented with G418 (Geneticin) and Hygromycin B respectively to a final concentration of 200 µg/ml. For selection of autotrophy, strains were growth onto synthetic medium supplemented with the required amino acids.

For loss of the *URA3* gene in *S. cerevisiae* or *ura4* gene in *Sz. pombe*, cells were grown in synthetic medium supplemented with FOA (5-fluoroorotic acid) to a final concentration of 0.1% (w/v). Yeasts cells carrying constructs cloned under the inducible *GAL1* promoter were grown on solid YPD medium or in liquid YP-Raff to suppress induction of the constructs or on solid or liquid

YPGAL for expression of the constructs. Table 2.3 gives detailed recipe of the media used (for growth of both yeasts and bacteria) in this study.

Table 2.1. List of *S. cerevisiae* strains used in this study**.** All strains were derived from and are isogenic to strain W303-1 (*ade2-1 ura3-1 his3-11,15 trp1-1 leu2-3,112 can1-100 rad5-535),* unless stated otherwise.

ID	Genotype	Source
CS1	*MATa*	Laboratory collection.
CS6	*MATα*	Laboratory collection.
CS44	*MATa top1Δ*::*kanMX*	Laboratory collection.
CS55	*MATa*/*MATα*	Laboratory collection.
CS74	*MATa pep4Δ*::*ADE2+*	Laboratory collection.
CS114	*MATa*/*MATα ade2-101/ade2-101 his3/his3-D200 leu2-3,112/ leu2-3,112 trp1-901/trp1-901 ura3-52/ura3-52 gal4Δ/gal4-452 gal80Δ/gal80-538 LYS2/lys2-801::GAL1UAS-GAL1TATA-HIS3 URA3::UASGAL1-LacZ, (met-)/URA3:: GAL4 17mers (X3)-CyC1TATA-LacZ*	Hybrigenics (not isogenic with W303).
CS1125	*MATa TAP-SLD5* (*kanMX*) *SEN1-9MYC* (*hphNT*) *pep4Δ*::*URA3+ ADE2+*	This study.
CS1126	*MATa SEN1-9MYC* (*hphNT*) *pep4Δ*::*URA3+ ADE2+*	This study.
CS1134	*MATa DPB2-TAP* (*kanMX*) *SEN1-9MYC* (*hphNT*) *pep4Δ*::*URA3+ ADE2+*	This study.
CS1187	*MATa TAP-SLD5* (*kanMX*) *SEN1-9MYC* (*hphNT*) *pep4Δ*::*URA3+ ADE2+ ctf4Δ*::*kanMX*	This study.
CS1217	*MATa TAP-SLD5* (*kanMX*) *SEN1-9MYC* (*hphNT*) *pep4Δ*::*URA3+ ADE2+ tof1Δ*::*HISMX*	This study.
CS1353	*MATa SEN1-TAP* (*kanMX*) *pep4Δ*::*ADE2+*	This study.
CS1403	*MATa POL12-TAP* (*kanMX*) *SEN1-9MYC* (*hphNT*) *pep4Δ*::*ADE2+*	This study.
CS1416	*MATa TAP-MCM3* (*kanMX*) *SEN1-9MYC* (*hphNT*) *pep4Δ*::*ADE2+*	This study.

CS1534	*MATa TAP-SLD5* (*kanMX*) *SEN1-9MYC* (*hphNT*) *pep4Δ*::*URA3+ ADE2+ mrc1Δ*::*hphNT*	This study.
CS1561	*MATa TAP-SLD5* (*kanMX*) *SEN1-9MYC* (*hphNT*) *pep4Δ*::*URA3+ ADE2+ csm3Δ*::*hphNT*	This study.
CS1671	*MATa POL12-TAP* (*kanMX*) *SEN1-9MYC* (*hphNT*) *pep4Δ*::*ADE2+ ctf4Δ*::*kanMX*	This study.
CS1676	*MATa TAP-SLD5* (*kanMX*) *SEN1-9MYC* (*hphNT*) *pep4Δ*::*URA3+ ADE2+ top1Δ*::*kanMX*	This study.
CS1711	*MATa TAP-MCM3* (*kanMX*) *GAL1-3HA-ø* (*LEU2+*) *pep4Δ*::*ADE2+*	This study.
CS1714	*MATa TAP-MCM3* (*kanMX*) *GAL1-3HA-SEN1 (2-931)* (*LEU2+*) *pep4Δ*::*ADE2+*	This study.
CS1852	*MATa GAL1-TAP-ø* (*LEU2+*) *pep4Δ*::*ADE2+*	This study.
CS1933	*MATa GAL1-TAP-SEN1 (1095-2231)* (*LEU2+*) *pep4Δ*::*ADE2+*	This study.
CS1941	*MATa GAL1-TAP-SEN1 (2-2231)* (*LEU2+*) *pep4Δ*::*ADE2+*	This study.
CS1942	*MATa GAL1-TAP-SEN1 (2-1901)* (*LEU2+*) *pep4Δ*::*ADE2+*	This study.
CS1943	*MATa GAL1-TAP-SEN1 (931-2231)* (*LEU2+*) *pep4Δ*::*ADE2+*	This study.
CS1956	*MATa GAL1-TAP-SEN1 (2-1103)* (*LEU2+*) *pep4Δ*::*ADE2+*	This study.
CS1957	*MATa GAL1-TAP-SEN1 (2-931)* (*LEU2+*) *pep4Δ*::*ADE2+*	This study.
CS2030	*MATa TAP-MCM3* (*kanMX*) *GAL1-3HA-SEN1 (2-622)* (*LEU2+*) *pep4Δ*::*ADE2+*	This study.
CS2032	*MATa TAP-MCM3* (*kanMX*) *GAL1-3HA-SEN1 (410- 931)* (*LEU2+*) *pep4Δ*::*ADE2+*	This study.
CS2056	*MATα td-MYC-sen1-1* (*K.l.TRP1+*) *GAL1-UBR1* (*HISMX*) *GAL1-TAP- ø* (*LEU2+*)	This study.
CS2058	*MATα td-MYC-sen1-1* (*K.l.TRP1+*) *GAL1-UBR1* (*HISMX*) *GAL1-TAP- SEN1 (2-931)* (*LEU2+*)	This study.
CS2061	*MATα td-MYC-sen1-1* (*K.l.TRP1+*) *GAL1-UBR1* (*HISMX*) *GAL1-TAP- SEN1 (2-1901)* (*LEU2+*)	This study.

CS2062	*MATα td-MYC-sen1-1* (*K.l.TRP1+*) *GAL1-UBR1* (*HISMX*) *GAL1-TAP- SEN1 (1095-2231)* (*LEU2+*)	This study.
CS2145	*MATa TAP-MCM3* (*kanMX*) *GAL1-3HA-SEN1 (410- 913)* (*LEU2+*) *pep4Δ::ADE2+*	This study.
CS2146	*MATa TAP-MCM3* (*kanMX*) *GAL1-3HA-SEN1 (410- 761)* (*LEU2+*) *pep4Δ::ADE2+*	This study.
CS2147	*MATa TAP-MCM3* (*kanMX*) *GAL1-3HA-SEN1 (410- 501)* (*LEU2+*) *pep4Δ::ADE2+*	This study.
CS2148	*MATa TAP-MCM3* (*kanMX*) *GAL1-3HA-SEN1 (501- 931)* (*LEU2+*) *pep4Δ::ADE2+*	This study.
CS2149	*MATa TAP-MCM3* (*kanMX*) *GAL1-3HA-SEN1 (761- 931)* (*LEU2+*) *pep4Δ::ADE2+*	This study.
CS2150	*MATa TAP-MCM3* (*kanMX*) *GAL1-3HA-SEN1 (622-931)* (*LEU2+*) *pep4Δ::ADE2+*	This study.
CS2184	*MATα td-MYC-sen1-1* (*K.l.TRP1+*) *GAL1-UBR1* (*HISMX*) *GAL1-TAP- SEN1 (2-1103)* (*LEU2+*)	This study.
CS2188	*MATα td-MYC-sen1-1* (*K.l.TRP1+*) *GAL1-UBR1* (*HISMX*) *GAL1-TAP- SEN1 (2-2231)* (*LEU2+*)	This study.
CS2276	*MATa hrp1Δ::kanMX*	Laboratory collection.
CS2334	*MATa/MATα SEN1/SEN1 (1-410, 411-2231 Δ)::URA3-CP*	This study.
CS2335	*MATa/MATα SEN1/SEN1 (1-622, 623-2231 Δ)::URA3-CP*	This study.
CS2403	*MATa/MATα SEN1/SEN1 (SEN1promoter-930 Δ)::URA3-CP*	This study.
CS2404	*MATa/MATα SEN1/SEN1 (SEN1promoter-912 Δ)::URA3-CP*	This study.
CS2451	*MATα td-MYC-sen1-1* (*K.l.TRP1+*) *GAL1-UBR1* (*HISMX*) *GAL1-TAP- SEN1 (931-2231)* (*LEU2+*)	This study.
CS2457	*MATa/MATα SEN1/SEN1 (913-2231)* (*HISMX*)	This study.
CS2458	*MATa/MATα SEN1/SEN1 (931-2231)* (*HISMX*)	This study.
CS2582	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (931-2231)* (*LEU2+*)	This study.

CS2584	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231)* (*LEU2+*)	This study.
CS2586	*MATa sen1Δ::URA3-CP SEN1-3HA-SEN1 (2-2231)* (*LEU2+*)	This study.
CS2603	*MATa GAL1-TAP-SEN1 (2-931)* (*LEU2+*) *pep4Δ::ADE2+ ctf4Δ::kanMX*	This study.
CS2607	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) W773A E774A W777A* (*LEU2+*)	This study.
CS2609	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) D850A E851G V852A L853G L854A* (*LEU2+*)	This study.
CS2617	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) V746G D747G P748G I749G* (*LEU2+*)	This study.
CS2623	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) L656A S657A K658A I659A L660* (*LEU2+*)	This study.
CS2636	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) L656A S657A K658A I659A L660A NRD1-9MYC* (*HIS3MX*) *pep4Δ:: ADE2+ TAP-MCM3* (*kanMX*)	This study.
CS2638	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231)W773A E774A W777A NRD1-9MYC* (*HIS3MX*) *pep4Δ:: ADE2+ TAP-MCM3* (*kanMX*)	This study.
CS2640	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) D850A E851G V852A L853G L854A NRD1-9MYC* (*HIS3MX*) *pep4Δ:: ADE2+ TAP-MCM3* (*kanMX*)	This study.
CS2642	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) V746G D747G P748G I749G NRD1-9MYC* (*HIS3MX*) *pep4Δ:: ADE2+ TAP-MCM3* (*kanMX*)	This study.
CS2656	*MATα sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231)* (*LEU2+*) *top1Δ::kanMX*	This study.
CS2659	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) L656A S657A K658A I659A L660* (*LEU2+*) *top1Δ::kanMX*	This study.

CS2661	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) W773A E774A W777A* (*LEU2+*) *top1Δ::kanMX*	This study.
CS2668	*MATα sen1Δ::URA3-CP ACT1-3HA-SEN1 (931-2231)* (*LEU2+*) *top1Δ::kanMX*	This study.
CS2669	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) NRD1-9MYC* (*HIS3MX*) *pep4Δ:: ADE2+* TAP-MCM3 (*kanMX*)	This study.
CS2670	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) NRD1-9MYC* (*HIS3MX*) *pep4Δ:: ADE2+*	This study.
CS2696	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) W773A E774A W777A hrp1Δ::kanMX*	This study.
CS2702	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) L656A S657A K658A I659A L660* (*LEU2+*) *hrp1Δ::kanMX*	This study.
CS2729	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) D850A E851G V852A L853G L854A* (*LEU2+*) *hrp1Δ::kanMX*	This study.
CS2734	*MATa rnh1Δ:: hphNT rnh201Δ::HISMX*	This study.
CS2735	*MATα rnh1Δ:: hphNT rnh201Δ::HISMX*	This study.
CS2736	*MATa rnh1Δ:: hphNT rnh201Δ::HISMX sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) W773A E774A W777A* (*LEU2+*)	This study.
CS2738	*MATa rnh1Δ:: hphNT rnh201Δ::HIS3MX sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) D850A E851G V852A L853G L854A* (*LEU2+*)	This study.
CS2791	*MATa td-sld3-7* (*kanMX*) *GAL1-UBR1* (*HIS3MX*) *GAL1-TAP-SEN1 (2-931)* (*LEU2+*) *pep4Δ:: ADE2+*	This study.
CS2798	*MATa CTF4 (1-367)-9MYC* (*kanMX*) *pep4Δ:: ADE2+*	This study.
CS2799	*MATa CTF4 (1-367)-9MYC* (*kanMX*) *pep4Δ:: ADE2+ GAL1-TAP-SEN1 (2-931)* (*LEU2+*)	This study.
CS2800	*MATa CTF4-9MYC* (*kanMX*) *pep4Δ:: ADE2+ GAL1-TAP-SEN1 (2-931)* (*LEU2+*)	This study.
CS2801	*MATa CTF4-9MYC* (*kanMX*) *pep4Δ:: ADE2+*	This study.

CS2802	*MATa 9MYC-CTF4* (*kanMX*) *pep4Δ*:: *ADE2+ GAL1-TAP-SEN1 (2-931)* (*LEU2+*)	This study.
CS2803	*MATa 9MYC-CTF4* (*kanMX*) *pep4Δ*:: *ADE2+*	This study.
CS2804	*MATa 9MYC-CTF4 (351- 927)* (*kanMX*) *pep4Δ*:: *ADE2+ GAL1-TAP-SEN1 (2-931)* (*LEU2+*)	This study.
CS2805	*MATa 9MYC-CTF4 (351- 927)* (*kanMX*) *pep4Δ*:: *ADE2+*	This study.
CS2806	*MATa CTF4 (1- 841)-9MYC* (*kanMX*) *pep4Δ*:: *ADE2+ GAL1-TAP-SEN1 (2-931)* (*LEU2+*)	This study.
CS2903	*MATa td-sld3-7* (*kanMX*) *GAL1-UBR1* (*HIS3MX*) *GAL1-TAP-SEN1 (2-931)* (*LEU2+*) (*kanMX*) *pep4Δ*:: *ADE2+ ctf4Δ*:: *kanMX*	This study.
CS2853	*MATa SEN1-TAP* (*kanMX*) *pep4Δ*:: *ADE2+*	This study.
CS2854	*MATa SEN1* (*W773A E774A W777A*)*-TAP* (*kanMX*) *pep4Δ*:: *ADE2+*	This study.

Table 2.2. List of *Sz. pombe* strains used in this study. All strains were derived either from the homothallic (self-fertile) 968 h^{90} strain or from the heterothallic (non-self-fertile) strains 972 h^{-} and 975 h^{+} originally isolated by Urs Leupold.

ID	Genotype	Source
JZ1	*h^{90} ade6-M210 leu1-32 ura4-D18*	Laboratory collection.
SV46	*h^{90} ade6-216 leu1-32 ura4D-18*	Laboratory collection.
FMA3	*h^{90} ade6-216 leu1-32 ura4D-18 loxP-ura4-loxP-pol1*	This study.
FMA4	*h^{90} ade6-216 leu1-32 ura4D-18 loxP-pol1*	This study.
FMA8	*h^{90} ade6-216 leu1-32 ura4D-18 loxP-pol1-ura4-loxM3*	This study.
FMA53	*h^{90LEU2} ade6+ leu1-32 ura4+ spp1-GFP:: kanR*	This study. Derived from P903. (Yang et al., 2005)
FMA54	*h^{90LEU2} ade6+ leu1-32 ura4+ psf2-YFP:: kanR*	This study. Derived from P1411. (Yang et al., 2005)
FMA63	*h^{90LEU2} ade6+ leu1-32 ura4+ rpl42P56Q^{cyhR}*	Laboratory collection.
FMA69	*h^{90LEU2} ade6-M210 leu1-32 ura4+ pol1-ts13 (ts) his3-D1*	This study. Derived from DBts131. (Bhaumik and Wang, 1998)
FMA70	*h^{90LEU2} ade6+ leu1-32 ura4+ pol1-H4 (ts)*	This study. Derived from SP262. (Murakami and Okayama, 1995)
FMA71	*h^{90LEU2} ade6-M216 leu1-32 ura4-D18 swi7-1*	This study.
FMA72	*h^{90} ade6-216 leu1-32 ura4D-1 loxP-pol1-loxM3*	This study.

FMA73	*h^{90} ade6-216 leu1-32 ura4D-18 loxP-pol1G1116S-loxM3*	This study.
FMA75	*h^{90} ade6-216 leu1-32 ura4D-18 loxP-pol1G1116Q-loxM3*	This study.
FMA76	*h^{90} ade6-216 leu1-32 ura4D-18 loxP-pol1G1116D-loxM3*	This study.
FMA78	*h^{90} ade6-216 leu1-32 ura4D-18 loxP-pol1G1116E-loxM3*	This study.
FMA85	*h^{90} ade6+ leu1-32 ura4-D18 loxP-9His-5FLAG-pol1-loxM3*	This study.
FMA86	*h^{90} ade6+ leu1-32 ura4-D18 loxP-9His-5FLAG-swi7-1-loxM3*	This study.
FMA93	*h^{90} ade6-M210 leu1-32 ura4-D18 swi1Δ::ura4+*	Laboratory collection.
FMA95	*h^{90} ade6-M210 leu1-32 ura4-D18 swi3Δ::kanMX*	Laboratory collection.
FMA131	*h^{90LEU2} ade6+ leu1-32 ura4+ pol1-1 (ts)*	This study. Derived from SP246. (D'Urso et al., 1995)
FMA132	*h^{90LEU2} ade6+ leu1-32 ura4+ pol1-1 (ts)*	This study. Derived from SP246. (D'Urso et al., 1995)
FMA146	*h^{90LEU2} ade6+ leu1-32 ura4+ mcl1Δ^{Toda}::kanMX*	This study. Derived from YAP12. (Mamnun et al., 2006)
FMA155	*h^{90LEU2} ade6-M216 leu1-32 ura4+ mcl1Δ^{Bioneer}::kanMX*	This study. Parental strain: knockout collection.

Table 2.3. Recipe of media used in this study for yeast and bacterial growth.

Medium	Recipe
YEA (Rich medium for *Sz. pombe*)	0.5% (w/v) yeast extract (Difco) 0.0225% (w/v) adenine 3% (w/v) glucose Optional: 2% (w/v) Bacto agar (Difco) for solid medium.
YEA + G418	0.5% (w/v) yeast extract (Difco) 0.0225% (w/v) adenine 3% (w/v) glucose 0.2 mg/ml Geneticin (G418) (Invitrogen) 2% (w/v) Bacto agar (Difco)
YEA + HU	0.5% (w/v) yeast extract (Difco) 0.0225% (w/v) adenine 3% (w/v) glucose 8 mM HU (Sigma) 2% (w/v) Bacto agar (Difco)
YEA + MMS	0.5% (w/v) yeast extract (Difco) 0.0225% (w/v) adenine 3% (w/v) glucose 0.0075% (v/v) (Sigma) 2% (w/v) Bacto agar (Difco)
YE (Rich medium for *Sz. pombe*)	0.5% (w/v) yeast extract (Difco) 3% (w/v) glucose Optional: 2% (w/v) Bacto agar (Difco) for solid medium.
PMA+ (Sporulation medium for *Sz. pombe*)	0.3% (w/v) KH phthalate 0.18% (w/v) Na_2HPO_4 0.5% (w/v) NH_4Cl 1% (w/v) glucose 0.2% (w/v) Kaiser SC 0.1% (v/v) vitamin mix (see below) 0.1% (v/v) minerals mix (see below) Optional: 2% (w/v) Bacto agar (Difco) for solid medium. pH adjusted to 6.0
Selective medium (For both *S. cerevisiae* and *Sz. pombe*)	0.17% (w/v) yeast nitrogen base (Difco) 0.5% (w/v) NH_4SO_4 0.2% (w/v) glucose 0.2% (w/v) Kaiser SC single Drop-out (Formedium). Optional: 2% (w/v) Bacto agar (Difco) for solid medium. pH adjusted to 6.0
YPD (Rich medium for *S. cerevisiae*).	1% (w/v) yeast extract (Bacto) 2% (w/v) peptone (Oxoid) 2% (w/v) glucose Optional: 2% (w/v) Bacto agar (Difco) for solid medium.
YPD + G418	1% (w/v) yeast extract (Bacto)

	2% (w/v) peptone (Oxoid) 2% (w/v) glucose 0.2 mg/ml Geneticin (G418) (Invitrogen) 2% (w/v) Bacto agar (Difco)
YPD + HygromycinB	1% (w/v) yeast extract (Bacto) 2% (w/v) peptone (Oxoid) 2% (w/v) glucose 0.3 mg/ml HygromycinB (Hygromycin B Gold™, InvivoGen) 2% (w/v) Bacto agar (Difco)
YPD + HU	1% (w/v) yeast extract (Bacto) 2% (w/v) peptone (Oxoid) 2% (w/v) glucose 50-100 mM HU (Sigma) 2% (w/v) Bacto agar (Difco)
YPD + MMS	1% (w/v) yeast extract (Bacto) 2% (w/v) peptone (Oxoid) 2% (w/v) glucose 0.005- 0.0075% (v/v) (Sigma) 2% (w/v) Bacto agar (Difco)
RSM (Sporulation medium for *S. cerevesiae*)	0.25% (w/v) yeast extract (Bacto) 1.5% (w/v) $K(C_2H_3CO_2)$ 0.1% (w/v) glucose 2.5% (v/v) amino acid mix
FOA plates (select against uracil autotrophy) (For both *S. cerevisiae* and *Sz. pombe*)	0.7% (w/v) yeast nitrogen base (w/ NH_4Cl) 2% (w/v) glucose 0.1% (w/v) 5-FOA 0.005% (w/v) uracil 0.225% (w/v) adenine 0.225% (w/v) leucine 0.225% (w/v) histidine 2% agar
Amino acid mix	0.4% (w/v) adenine 0.2% (w/v) arginine 0.4% (w/v) histidine 0.2% (w/v) leucine 0.2% (w/v) lysine 0.2% (w/v) methionine 1% (w/v) phenylalanine 0.2% (w/w) tryptophan 0.08% (w/w) tyrosine
Vitamins mix	0.1% (w/v) pantothenic acid 1% (w/v) nicotinic acid 1% (w/v) inositol 0.001% (w/v) biotin
Minerals mix	5% (w/v) boric acid 4% (w/v) $MnSO_4$

	4% (w/v) $ZnSO_4.7H_2O$ 2% (w/v) $FeCl_2.6H_2O$ 0.04% (w/v) molybdic acid 0.1% (w/v) KI 0.04% (w/v) $CuSO_4.5H_2O$ 1% (w/v) citric acid
Luria-Bertani (LB) Broth	1% (w/v) bacto-tryptone 0.5% (w/v) yeast extract (Bacto) 1% (w/) NaCl pH adjusted to 7.0 Optional:2% (w/v) agar Optional: Supplemented with either 100 µg/ml ampicillin or 50 µg/ml kanamycin.
NZY+ Broth	1% (w/v) NZ amine (casein hydrolysate) 0.5% yeast extract 0.5% NaCl pH adjusted to 7.5 12.5 mM $MgCl_2$ 12.5 mM $MgSO_4$ 0.4% glucose
SOC medium	2% (w/v) tryptone 0.5% (w/v) yeast extract 10 mM NaCl 2.5 mM KCl 10 mM MgCl2 10 mm MgSO4 Final pH 6.8-7.0 20 mM glucose (added prior to use)
RF1 medium	1.2% (w/v) rubidium chloride 0.99 % (w/v) $MnCl_2.4H_2O$ 3% (w/v) 1M potassium acetate pH 7.5 0.15% (w/v) $CaCl_2.2H_2O$ 1.5% (v/v) glycerol Final pH 5.8 with acetic acid.
RF2 medium	2% (v/v) 0.5M MOPS pH 6.8 0.12% (w/v) RbCl 10mM KCl 1.1% (w/v) $CaCl_2.2H_20$ 1.5% (v/v) glycerol Final pH 6.8 with NaOH.

2.1.2 CROSSING OF STRAINS

Cells were grown from 25% (v/v) glycerol suspensions onto non-selective plates (either YPD or YEA) and grown at the appropriate temperatures. For *S. cerevisiae* strains, two parents of opposite mating-types were mixed in a ratio of about 20:1 in 50 µl of sterile deionized water to homogeneity. 10 µl of the resulting cell suspension was then plated onto a non-selective plate and grown overnight at the appropriate temperature. This allows enough time for the two parent strains to mate to form diploids. The mixture was then plated onto a selective medium so as to select against the parental strain found in excess and grown overnight at the appropriate temperature. The mixture was then plated onto sporulation medium and grown until the cells sporulated (typically 3-5 days). The resulting asci were then treated with gluronidase from snails' gut (*Helix pomata*) and their tetrads were aligned by micro-manipulation onto non-selective plates and allowed to germinate. The tetrads were scored by replica-plating on different selective media.

Unlike *S. cerevisiae*, *Sz. pombe* cells prefer to be haploid and, upon diploidization, will undergo sporulation spontaneously. To cross strains, cells of compatible mating-type types were mixed onto sporulation medium. When crossing an h^+ with an h^- strain or two homothallic (switching or self-fertile) (h^{90}) strains, equal quantities of each parent was used. When crossing a homothallic strain with a heterothallic strain, however, the heterothallic (non-switching or non-self-fertile) strain was added in large excess (a ratio of about 20:1). This is because cells from homothallic strains would tend to mate with their sister or cousin cells by virtue of cell-type switching and proximity. The mixture was left on sporulation medium for at least three days at the appropriate temperature. This allows for the formation of transient diploids (zygote) that undergo meiosis before undergoing sporulation. Mating was judged successful when zygotic asci were detected microscopically. The zygotic asci were treated with glucoronidase overnight at 37°C. This treatment simultaneously kills vegetative cells and liberate spores from their asci. The supernatant was separated from the pellet containing cell debris and spores by centrifugation and the pellet was re-suspended in 30% (w/v) ethanol. This

inactivates any vegetative cells and helps to deter bacterial growth. 1/100 and 1/1000 dilutions of the pellets were then plated onto the appropriate selective media.

For tetrad dissection, we made use of the *ade6-M210* and *ade6-M216* markers. In brief, parents of compatible cell-types (see above) were mated onto sporulation medium. In addition, one parent carried the *ade6-M210* marker whilst the other parent carried the *ade6-M216* marker. As such, both parents are auxotrophs for adenine. However, the *ade6-M210* and *ade6-M216* alleles complement one another at an intragenic level within diploids. This enables isolation of stable diploids. The cell mixture was plated onto synthetic medium lacking adenine 1, 2 and 3 days after mixing the parental strains. The diploid colonies were then grown on fresh synthetic medium lacking adenine before being plated on sporulation medium. After sporulation, azygotic asci were left overnight at 16°C or 3-5 h at 37°C for their walls to be degraded. The liberated tetrads were then aligned by micro-manipulation onto non-selective plates and allowed to germinate. The tetrads were scored by replica-plating on different selective media.

2.1.3 DILUTION SPOTTING

Cells were grown from 25% (v/v) glycerol suspensions onto non-selective medium and grown for at least two days at the required temperature (up to four days for strains grown at 25°C). Cells were diluted in sterile d.H_2O water supplemented with 100 µg/ml of ampicillin to prevent growth of bacterial contaminants to a final concentration of 5 x10^6 cells/ml. This suspension was serially diluted ten-fold to produce suspensions of 5 x 10^5, 5 x 10^4 and 5 x10^3 cells/ml. 10 µl of each suspension was pipetted onto either non-selective or selective media and grown for up to 5 days at the required temperatures. The colonies were imaged daily using an Epson Perfection V700 Scanner.

2.1.4 TRANSFORMATION OF YEAST CELLS USING THE LITHIUM ACETATE METHOD

Cycling cells were grown to a density of 1 x 10^7 cells/ml. *S. cerevisiae* cells were washed with sterile d.H_2O and re-suspended in a solution of 0.1 M lithium acetate (pH 7.5), 1 M Tris-HCl (pH 7.5) to a final concentration of 2.0 x 10^9 cells/ml. To 100 µl of the cell suspension, 1-2 µg of target DNA (plasmid, PCR or restriction digest) and an excess (500 µg) of freshly denatured carrier ssDNA (extracted from salmon sperm) were added in a combined volume of 10 µl. To this mixture, 40% (w/v) PEG 4000 in 0.1 M lithium acetate (pH 7.5) 1 M Tris-HCl (pH 7.5) was added to a final concentration of 33.3% (w/v) PEG, vortexing to achieve homogeneity. The cell suspension was then incubated for 45 min at 25°C before addition of sterile DMSO to a final concentration of 10%. After addition of DMSO, the cell suspension was heat-shocked at 42°C for 15 min before rapid cooling on ice for 2 min. The suspension was then centrifuged and the supernatant discarded. If the transformation involved autotrophic markers, the cells were re-suspended in sterile d.H_2O and different quantities of the suspension were plated on the appropriate selective media. Had the transformation involved either the *kanMX* or *hphNT* markers, the cells were re-suspended in non-selective media and grown for a further 3 h at 25°C to allow for expression of the antibiotic before plating on the appropriate selective media.

For *Sz. pombe*, a similar method was used with some modifications. In brief, cycling *Sz. pombe* cells were grown to a density of 1 x 10^7 cells/ml, washed in sterile d.H_2O and re-suspended in a 0.1 M lithium acetate (pH 4.5) and dispensed in aliquots of 100 µl. The cells were incubated at 30°C for 2 h. Thereafter, 1 µg of target DNA in 15 µl of TE (pH 7.5) was added, vortexing to homogeneity. 50% (w/v) of PEG 4000 pre-warmed to 30°C was added to a final concentration of 35.8% (w/v). The mixture was incubated for another hour at 30°C. The mixture was them heat-shocked for 15 min at 43°C before rapid cooling on ice for 2 min. The suspension was then centrifuged with the supernatant discarded. The cells were re-suspended in non-selective media

and grown for a further 3 h at 30°C before plating on the appropriate selective media.

2.1.5 TESTING FOR THE CELL-TYPE OF HAPLOIDS

The mating/cell-type of *S. cerevisiae* cells was assessed by their reaction to exposure to the α-factor pheromone to a final concentration of 7.5 µg/ml for 2-3 h and by their ability to mate with tester *MATa* and *MATα* strains to form diploids capable of sporulating on the appropriate media.

The mating-type of *Sz. pombe* cells was also assessed by their ability to mate with tester strains (h^+ and h^-) to form diploids capable of sporulating on the appropriate media. Alternatively, the mating type was determined by colony PCR. In brief, single colonies of *Sz. pombe* were boiled in sterile d.H_2O for 7 min, releasing genomic DNA into the water. The water was used as template for PCR using a primer that is common to both *mat1-P* and *mat1-M* (NRT40) with primers specific to *mat1-P* (NRT41) and *mat1-M* (NRT42) respectively (Table 2.5). Strains harbouring the h^+ mating-type are characterized by a PCR product of 987 bases long whilst h^- strains are characterized by a product of 729 bases long. Diploids or homothallic strains are characterized by both products.

2.1.6 TESTING THE EFFICIENCY OF CELL-TYPE SWITCHING IN FISSION YEAST

Unlike *S. cerevisiae*, the spores of fission yeasts are characterized by the presence of an amylose-like polymer that stains with iodine. To determine whether cells were defective for cell-type switching, h^{90} or h^{90}*:: LEU2* cells freshly grown on non-selective media were first transferred onto sporulating malt-extract medium in such a way as to obtain isolated colonies. The cells were grown either for 4-5 days at 25°C for heat-sensitive strains or for 2-3 days at 33°C (until colonies become sufficiently large) before growth for an additional day at 30°C. This allows haploids to diploidize and to sporulate.

Importantly, each plate used contained two strains; a strain under investigation and a wildtype strain (JZ1) as control.

The plates containing the colonies were held then upside down over a dish containing iodine crystals at 25°C for 5-6 min. The iodine sublimes and the vapour stain the colonies. Colonies that can switch cell-type would stain darkly (dark purple/chocolate brown) whilst colonies that are defective for cell-type switching stain lightly or with a speckled phenotype. The colonies were then imaged at a magnification of X 5 using a Leica MZ FL Stereomicroscope using a top-light to illuminate the colonies.

2.1.7 HARVESTING OF YEAST STRAINS FOR IMMUNO-PRECIPITATIONS

MATa strains were grown from 25% (v/v) glycerol suspensions onto solid non-selective medium and grown for 1-3 days at the appropriate temperatures. Fresh YPD was inoculated with the cells and cycling cultures were grown to a cellular density of 0.7 x 10^7 cells/ml in volumes of either 250 ml or 1 l for dilute or concentrated samples respectively at 24°C. The cells were then arrested in G_1 by addition of the α-factor pheromone to a final concentration of 7.5 µg/ml for at least 3 h. The cells could then be harvested in G_1 if so required. Otherwise, the arrested cells were washed twice with fresh YPD (lacking α-factor) and released in S phase at 24°C. Cells were harvested 30 min (S phase) or 60-65 min (G_2) after release from G_1-arrest. To ensure the cells were harvested in the correct phase of the cell cycle, 1 ml of cells were fixed in 70% (v/v) ethanol and stored at 4°C for FACS analysis.

To harvest the cultures, cells were first washed in 20 mM Hepes-KOH (pH 7.9) (Sigma) and then washed in a solution of 100 mM Hepes-KOH (pH 7.9) (Sigma), 50 mM potassium acetate (Fischer), 10 mM magnesium acetate (Sigma) and 2 mM EDTA-KOH at 4°C.

For dilute samples, 1.75 x 10^9 cells were then re-suspended in a solution of 100 mM Hepes-KOH (pH 7.9), 50 mM potassium acetate, 10 mM magnesium acetate and 2 mM EDTA-KOH supplemented with 2 mM glycerophosphate

(Johnson Matthey), 2 mM sodium fluoride (Fischer), 1 mM DTT, 1% (v/v) Sigma protease inhibitor cocktail (for fungal and yeast extracts, Sigma) and 0.24% (w/v) EDTA-free Complete Protease Inhibitor Cocktail (Roche) so that the ratio of the wet mass of the cells to the final mass of the suspension was 1:4. For concentrated samples, 7 x 10^9 cells were also re-suspended in the same solution but so that the ratio of the wet mass of the cells to the final mass of the suspension was 4:5 and the final concentrations of the protease and phosphatase inhibitors were increased 1.2-fold. The samples were maintained at 4°C before careful pipetting into liquid nitrogen. This freezes the cells, completing harvesting. The frozen cells were then kept at −80°C until use for immuno-precipitation. After harvesting, left-over cells at 4°C were fixed in 70% (v/v) ethanol for FACS analysis.

When using strains harbouring degron alleles (also carrying the *GAL1-UBR1* allele), cycling cells were grown to a density of 0.7 x 10^7 cells/ml at 25°C in 250 ml of YP-Raff and arrested in G_1 by addition of the α-factor pheromone to a final concentration of 7.5 μg/ml for at least 3 h. Before release in S phase, however, the cells were shifted first to YPGAL medium for 35 min to induce expression of the E3 ligase Ubr1 protein and then to 37°C for 1 h to allow for degradation of the degron-tagged proteins. Cells were then released in S phase. Harvesting was carried out 20 min after release. Likewise, when using slow-growing strains such as strains harbouring the *ctf4Δ* allele, cells were grown at 30°C and were harvested 20 min after release from G_1 arrest.

For cells harbouring constructs under control of the strong, inducible *GAL1* promoter, cultures were grown in YP-Raff or YPGAL. When using YP-Raff, cultures were arrested in G_1 by addition of α-factor for 3 h before the medium was substituted with YPGAL already supplemented with the pheromone. Cells were maintained in YPGAL for 35 min to allow for expression of constructs under control of the *GAL1* promoter before release in fresh YPGAL at 24°C. For experiments where YPGAL was used exclusively, the cells could be released in fresh YPGAL immediately upon reaching G_1 synchronycity, similar to cultures using only YPD.

2.2 *E. COLI* SPECIFIC METHODS

2.2.1 PREPARATION OF CHEMICALLY COMPETENT CELLS

DH5α cells were woken from 25% (v/v) glycerol suspensions on solid LB medium and grown for ~ 16 h at 37°C until sufficiently big colonies were obtained. 5-10 individual colonies were used to inoculate a fresh aliquot of 10 ml LB and the culture was grown at 37°C overnight. A portion of this inoculum was diluted in 100 ml of fresh LB medium that was incubated at 37°C with shaking until an OD_{600} of 0.5 was obtained (midlog phase). The culture was then transferred to sterile tubes and cooled by placing on ice for 10 min. The sample was centrifuged at 2700 *g* and the supernatant was removed by careful decantation. The pellet was re-suspended in 33 ml of RF1 solution to homogeneity and left to stand on ice for a further 10 min. The sample was centrifuged at 2700 *g* for 10 min and the RF1 solution was removed by decantation. The pellet was then re-suspended in 8 ml of RF2 to homogeneity and left to stand on ice for 10 min. 500 µl of cell suspension was aliquoted in Eppendorf tubes that were immediately flash-frozen in liquid nitrogen. The tubes were kept at −80°C until required.

2.2.2 TRANSFORMATION OF *E. COLI* CELLS

Chemically-competent cells suspensions (DH5α cells prepared in-house or commercially-sourced XL Gold (Agilent) or BL21 Star™ (DE3) (Invitrogen)) were thawed from −80°C on ice for 5-10 min. If required, 100-200 µl of cell suspension were dispensed in chilled Eppendorfs to which plasmid DNA (unprocessed plasmid or derived from either mutagenesis or ligation reaction) was added. Gentle tapping ensures mixing and the cell suspensions were then left on ice for 30 min (1 h for commercially-sourced cells). The cells were then heat-shocked for 90 s (30 s for commercially-sourced cells), followed by rapid cooling on ice for 2 min. The suspensions were then immediately diluted in 250-1000 µl of SOC medium (NZY+ for XL Gold cells) pre-warmed to 37°C and grown at 37°C with shaking for 1 h. This allows expression of the

appropriate antibiotics-resistance genes. The cells were then plated on the appropriate selective media at different dilutions and grown for ~ 16 h at 37°C.

2.3 MOLECULAR BIOLOGY

2.3.1 GENOTYPING BY PCR

To tag or delete genes, heterologous DNA with homology to the gene of interest and carrying DNA corresponding to the appropriate markers (and also, where appropriate, DNA corresponding to protein tags) were constructed by PCR. Yeast strains were transformed with purified DNA using the lithium acetate method (Section 2.1.4) and successful transformants were screened by growth on selective media. To check whether the novel DNA was introduced at the correct loci, we adopted a PCR-based approach where we designed oligoes that align upstream and downstream of the target loci and oligoes that align with the marker cassettes used. These oligoes were used to screen for correct transformants by colony-PCR or by PCR using genomic DNA extracted from those transformants.

2.3.2 CLONING OF INDUCIBLE CONSTRUCTS OF *SEN1*

To clone inducible fragments of *SEN1*, we used the integrative pRS305 plasmid. pRS305 is a derivative of pBLUESCRIPT, with a *LEU2* marker and a multiple cloning site (MCS) cloned in (Sikorski and Hieter, 1989).

We first amplified DNA encoding the *GAL1* promoter followed by repeats of the haemagglutinin tag from pYM-N24 (pCS11 in our lab) (Janke et al., 2004) using oligoes CS355 and CS356. The PCR product was designed with a NotI restriction site at its 5'-end and a SalI restriction site at its 3' end. The PCR product was introduced into the pRS305 plasmid by virtue of these sites to produce the pCS25 plasmid (pRS305-*GAL1*-3HA). Fragments of *SEN1* were amplified from genomic DNA extracted from strain CS1 (wildtype) to synthesize products with SalI sites at their 5' end and XhoI sites at their 3' ends. The fragments were then cloned in plasmid pCS25 and analysed by sequencing.

To tag constructs with the TAP tag instead, we used a derivative of pRS305, pKL935 that already had a *GAL1* promoter cloned between the NotI and XBaI sites. DNA encoding the NTAP2 tag followed by a GAGAGAGAGA linker was amplified from pKL205 using oligoes CS446 and CS447. The product was designed with a XbaI and SalI sites at its 5' and 3' ends respectively enabling cloning into pKL935 to give the pCS14 plasmid (pRS305-GAL1-NTAP2). Fragments of *SEN1* were amplified from genomic DNA extracted from strain CS1 (wildtype) to synthesize products with SalI sites at their 5' end and PspXI sites at their 3' ends. The fragments were then cloned in plasmid pCS14 and analysed by sequencing.

Derivatives of pCS14 and pCS25 were linearized with the XcmI enzyme that cuts within the *LEU2* marker and the purified digests were transformed into diploids cells homozygous for the *leu2-3,112* mutation. Upon successful integration, the constructs reconstitute the *LEU2* marker making the cells autotrophic for leucine. Integration was also checked by extracting genomic DNA from clones and using this DNA to perform diagnostic PCRs using oligoes CS28, CS29, CS32 and CS33. The clones were also screened by induction of protein expression in YPGAL and analysed by Western blotting. Haploids from clones adjudged to have the correct insert were obtained by tetrad dissection.

2.3.3 PROMOTER SWITCHING FROM *GAL1* TO *ACT1* OR *SEN1*

To produce non-inducible constructs of *SEN1*, we decided to substitute the *GAL1* promoter from pCS25-derivatives with either the *SEN1* promoter or the strong *ACT1* promoter. To introduce these promoters, the *GAL1* promoter was first excised from the pCS25-derivatives using the NotI and SpeI endonucleases. The *ACT1* promoter was amplified from genomic DNA using the oligoes CS953 and CS954 with a product encoding a NotI site at its 5' end and a SpeI site at its 3' end. Likewise, the *SEN1* promoter was amplified from genomic DNA using the oligoes CS955 and CS956 with a product encoding a NotI site at its 5' end and a SpeI site at its 3' end. The *ACT1* and *SEN1* promoters were then cloned into derivatives of the pCS25 plasmid. These plasmids were then linearized and introduced into diploids through transformation. The clones were analysed as described in Section 2.3.2.

2.3.4 CREATING *SEN1* POINT MUTANTS BY SITE-DIRECTED MUTAGENESIS

The plasmid pCS120 was obtained by substituting the *ACT1* promoter for the *GAL1* promoter as described in Section 2.3.3. To obtain novel alleles of *SEN1*, we first identified conserved residues with the armadillo-repeat motifs within the Sen1 N-terminal domain that were predicted to be on the surface of the super-helix and that were conserved. We then mutated the bases encoding for these residues using an *in vitro* site-directed mutagenesis protocol (QuikChange Lightning, Qiagen). In brief, a pair of complimentary oligoes with the desired mutation was used to amplify the target DNA, here pCS120. The kit uses a blend that includes the high-fidelity *Pfu* DNA polymerase, reducing the incidence of unwanted mutations. Whilst PCR products are usually linear, the newly synthesized DNA is circular by virtue of being attached to template circular DNA and because enzymes within the blend seal nicks.

After the amplification reaction, the product DNA is present in large excess compared to its template. Moreover, the template differs from the product DNA

by virtue of bacterial methylation of cytosine bases. As such, methylated and hemi-methylated DNA was digested using the DpnI nuclease. The DNA was then transformed without purification into XL-10 Gold Ultracompetent (Qiagen) cells to amplify the plasmids, according to the manufacturer's guidelines. Plasmids were isolated from single colonies and analysed by sequencing. Plasmids with the correct mutations were linearized with XcmI and transformed in diploid cells as described in Section 2.3.2, except that *SEN1*/*sen1*Δ strains were used instead. Clones with the correct integration were placed on sporulation media and their tetrads were dissected. Haploids autotrophic for both uracil and leucine carried only the mutant copy of *SEN1* albeit under control of the *ACT1* promoter cloned at an ectopic locus.

2.3.5 Generating alleles of *SEN1* lacking the N-terminus or with mini-truncations within their N-termini at their genomic locus

To create the *SEN1 (913-2231)* allele, diploids were transformed with products amplified from plasmid pKL506 using oligoes CS898 and CS818. This disrupts the *SEN1* allele upstream of its transcription start site to its 912th codon with a *URA3* gene. This produces a *SEN1*/*SEN1(1-912*Δ*, 913-2231)* diploid (CS2404). Oligoes CS898 and CS899 were used to amplify the *HISMX* marker from plasmid pKL230 and oligoes CS900 and CS901 were used to synthesize a PCR bridging the ATG of *SEN1* to the 913th codon of the *SEN1* gene was synthesized. The two PCRs were fused in another PCR reaction and this PCR was used to transform strain CS2404. To create the *SEN (931-2231)* allele, an identical approach was adopted where oligoes CS818 and CS901 were substituted with oligoes CS756 and CS902. This process is summarized in Fig 2.1.

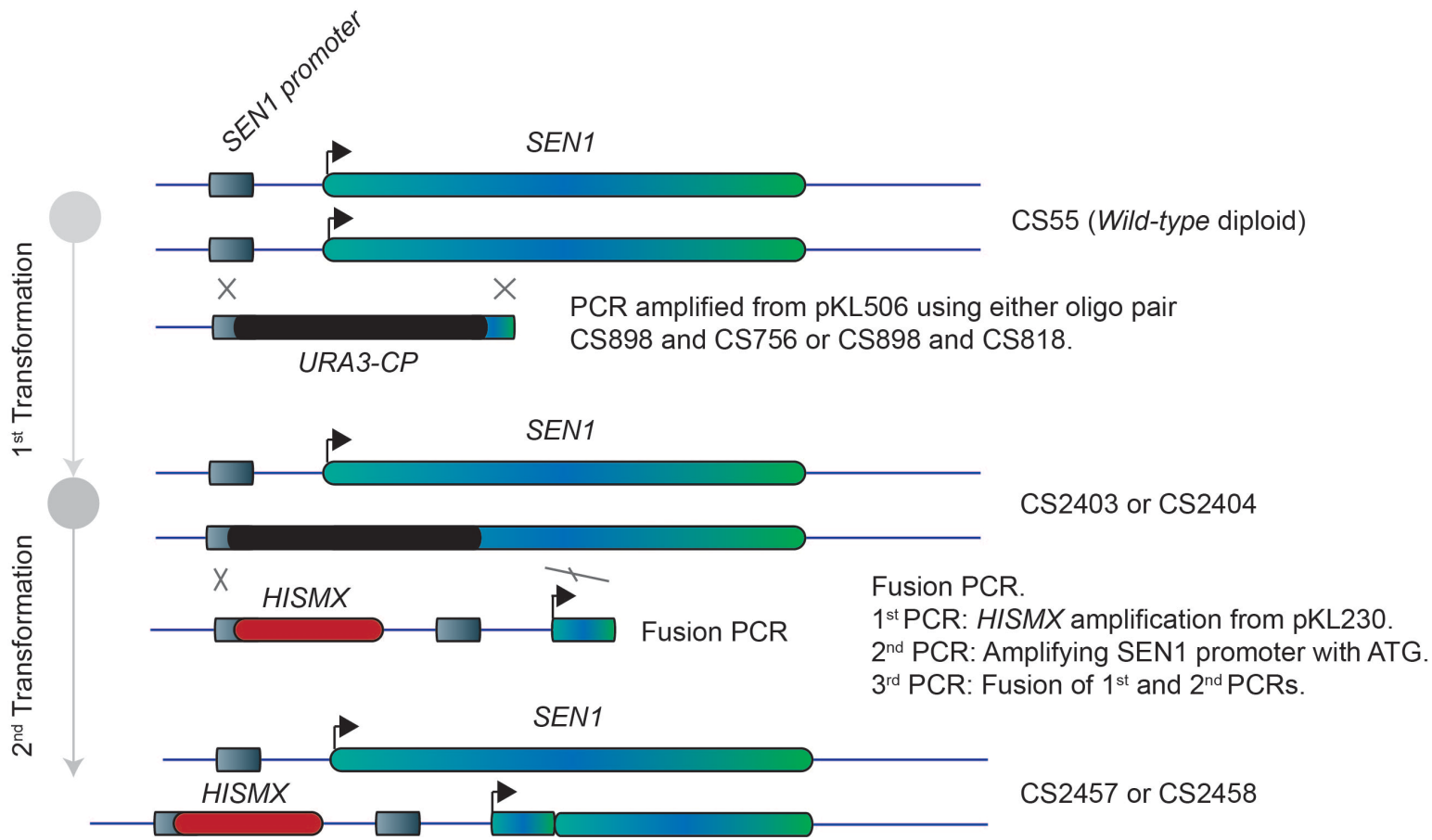


Figure 2.1. Schematic of the stratagem used to generate alleles of *SEN1* that lack the bases that encode for its N-terminal domain.

Because both the *SEN1 (913-2231)* and *SEN1 (931-2231)* were sick compared to their wildtype counterpart, we resorted to generate alleles of *SEN1* with mini-truncations within their N-termini. In order to do so, we first disrupted deleted the *SEN1* gene from either its 411th or its 623rd codon to its 2231st codon with the *URA3* marker. This was done by amplification of the pKL506 plasmid using oligoes CS755 and CS291, and CS817 and CS291 respectively, followed by transformation in a diploid. The resulting strains were heterozygous for *SEN1*: CS2334 (*SEN1*/*SEN1(1-410*, *411-2231*Δ)) and CS2335 (*SEN1*/*SEN1* (*1-622*, *623-2231*Δ)).

Genomic DNA was then extracted from strain CS1353 that harbours the *SEN1-TAP* allele. From this DNA, PCRs were generated using the CS291 oligo as the reverse primer and different forward primers to generate PCR products that bridge either the 410th or 622nd codon to either 913th or 931st codon. Moreover, we attempted to bridge the codons by several strategies. We either included no linkers, or a short linker (GSGAGSGAGSG), or a somewhat longer linker to bridge the gap (GSGAGSGSGGAGAGSGSGAGGSG) or alternatively we used the CIP (*C*tf4 Interacting Peptide) (GGSGGSIDNFDDILGEFEGAGSG) from Pol1. We did so as it is known that the solubility and stability of proteins can be compromised upon intra-domain tinkering.

The different PCRs were integrated into the appropriate diploid (either CS2334 or CS2335) and successful transformations were screened initially by growth on YPD supplemented with G418 and then by growth on FOA plates, indicating the loss of the *URA3* marker. The clones were then assessed by diagnostic PCRs and finally by Western blotting with strain CS1353 (that carries the *SEN1-TAP* allele) used as a control. For successful integration, not only would the TAP signal be detected but it would run at height of 220-250 kDa whilst Sen1-TAP would run at around 280 kDa. The clones that integrated correctly were placed on sporulation medium and their tetrad were dissected. This process is summarized in Fig 2.2.

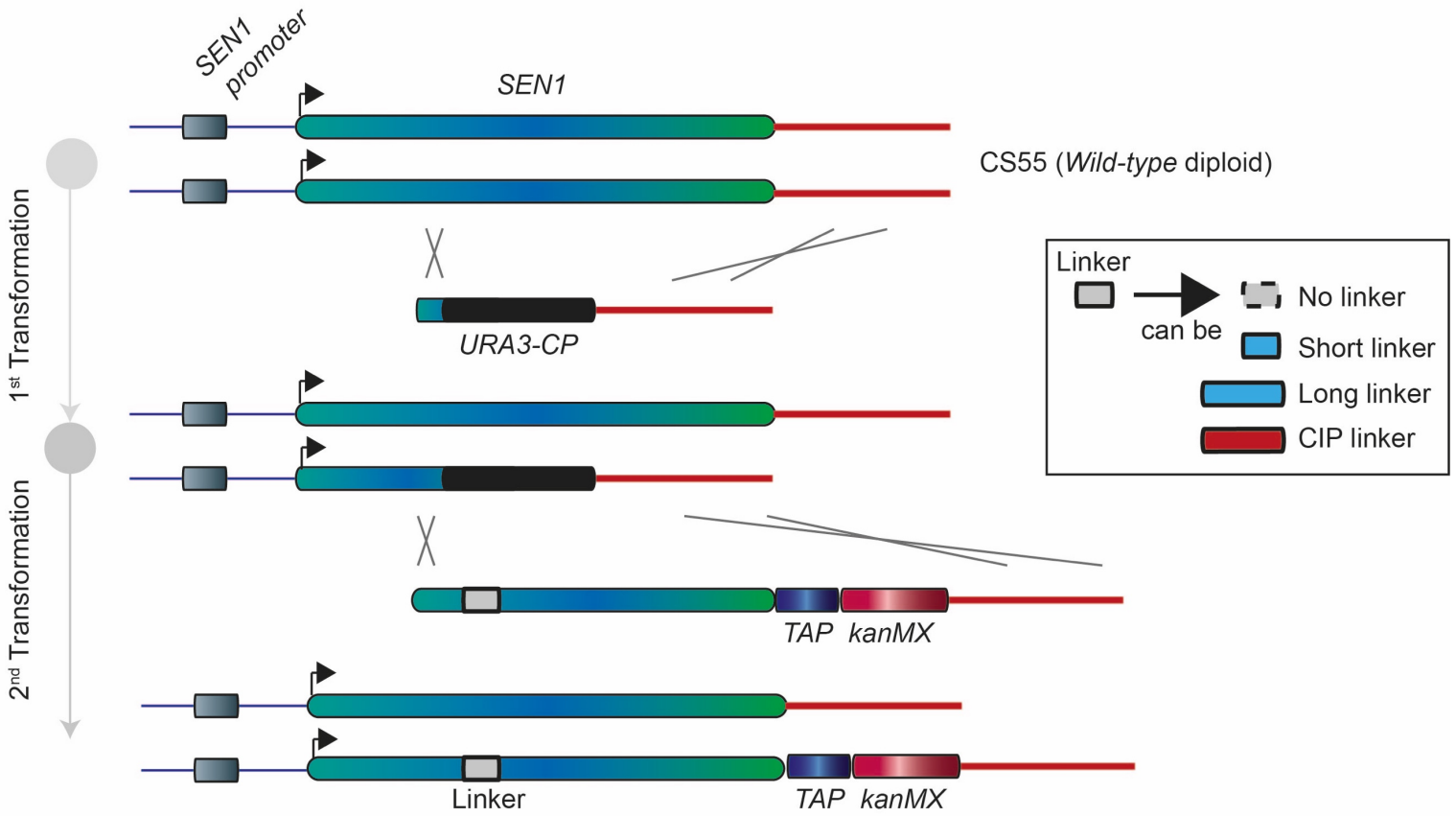


Figure 2.2. Schematic of the stratagem used to generate alleles of *SEN1* that encodes for variants of the protein with mini-truncations within their respective N-terminal domain.

2.3.6 TAGGING OF POL1 AT THE N-TERMINUS AND CREATING *POL1* MUTANTS BY SITE-DIRECTED MUTAGENESIS

Both for tagging and for creating novel *pol1* alleles in *Sz. pombe*, we used a cassette exchange protocol (Watson et al., 2008). In brief, a wildtype haploid cell (SV46) was used to construct a base strain. Initially, a heterologous DNA product was amplified from plasmid pAW41 using oligoes NRT14 and NRT15 and this DNA was transformed upstream of the *pol1* promoter using the lithium acetate method. Successful integration was checked by autotrophy for uracil and diagnostic PCRs of genomic DNA from transformants. The resulting strain (FMA3) was characterized by the l*oxP-ura4-loxP-pol1* allele. This strain was then transformed with the non-integrative plasmid pAW5 that expresses the *Cre*-recombinase as well as the *LEU2* marker. Successful transformation was adjudged by virtue of autotrophy for leucine. The clones were grown for 4 days at 33°C on selective medium lacking YNB (yeast nitrogen base) to allow expression of the *Cre*-recombinase under control of the *nmt1* promoter. The clones were then grown on FOA plates. Growth on FOA plates signals loss of the *ura4* marker, by recombination between the two *loxP* sites previously introduced upstream of the *pol1* gene. This leaves a *loxP* scar some 260 bases upstream of *pol1* whose presence was confirmed by diagnostic PCR. The corresponding strain (FMA4) was grown for several days on non-selective medium to lose the pAW5 plasmid and was then transformed by a heterologous DNA product amplified from plasmid pAW12 using oligoes NRT16 and NRT17. This introduces the *ura4* marker immediately downstream of the *pol1* gene along with the *loxM3* recombination site. Successful integration was checked by autotrophy for uracil and diagnostic PCRs of genomic DNA from transformants. The resulting strain (FMA8) was characterized by the l*oxP-pol1-ura4-loxM3* allele and was used as a base strain.

To create tagged or mutant alleles of *pol1*, the gene was first amplified from genomic DNA of a wildtype strain (JZ1) using oligoes NRT18 and NRT19 and inserted into the pAW8 plasmid between the *loxP* and *loxM3* sites to form the

pRA8 plasmid. The plasmid was checked for mutations by PCR. To tag *pol1*, we amplified two regions from plasmid pRA8. Using oligoes CS440 and CS441, a region immediately upstream of the *pol1*'s transcription start site was amplified whilst oligoes CS444 and CS445 were used to amplify a region immediately downstream of the transcription start. A final pair of oligoes (CS442 and CS443) was used to amplify a DNA region encoding for a 9XHIs-5XFLAG tag from plasmid pKL712. This third oligo was used to bridge the two earlier PCRs using fusion PCR. The resulting product was cloned into pRA8 using the SacI and XhoI sites to get the pRA22 plasmid. The newly cloned region was analysed by sequencing.

To generate novel alleles of *pol1*, plasmids pRA8 or pRA22 were used as templates for site-directed mutagenesis as described in Section 2.3.4. Plasmids pRA8, pRA22 and their derivatives were then transformed into the base strain (FMA8) using the lithium acetate method. Successful transformation was adjudged by virtue of autotrophy for leucine. The clones were grown for 4 days at 33°C on selective medium lacking YNB (yeast nitrogen base) to allow expression of the *Cre*-recombinase under control of the *nmt1* promoter. The clones were then grown on FOA plates. Growth on FOA plates signals loss of the *ura4* marker. The *loxP* and *loxM3* sites are not compatible and loss of *ura4* cannot happen by looping out the *pol1* gene from the genome. Instead loss of the *ura4* gene can be achieved by recombinant exchange of the cassette in the base strain with cassettes found on the pRA8, pRA22 or their derivatives. The resulting strains were analysed by diagnostic PCRs, sequencing and (where appropriate) Western blotting. Figure 2.3 summarizes this process.

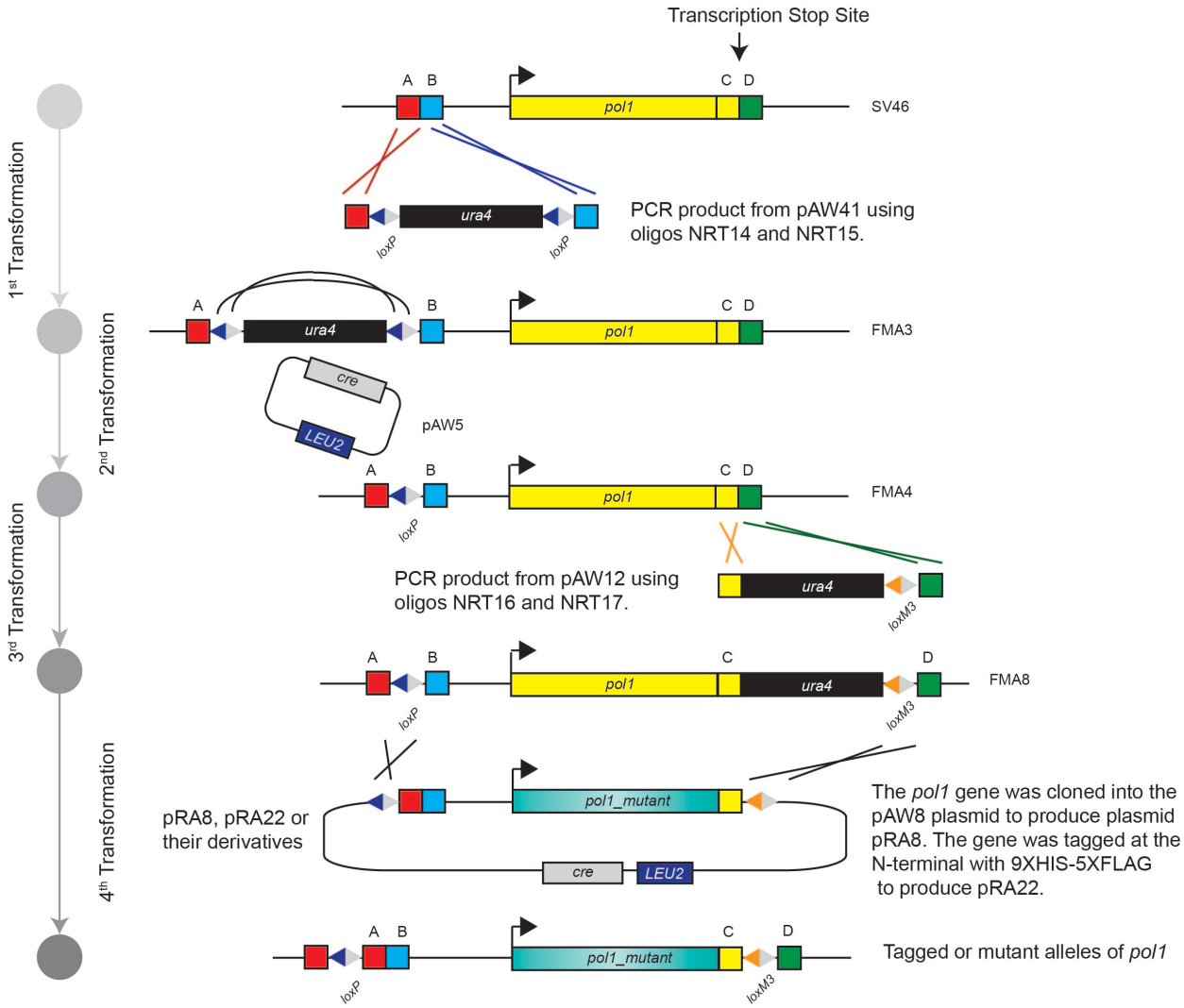


Figure 2.3. Schematic of the stratagem used to generate novel and tagged alleles of *pol1* in *Sz. pombe.*

2.3.7 LIST OF PLASMIDS AND OLIGOES USED IN PRESENT STUDY

Table 2.4. List of plasmids used.

Plasmid Name	Insert	Backbone	Use	Source
Plasmids used for the *Sz. pombe* study				
pAW5	Empty	pAW5	Cloning of *loxp-pol1-loxM3* and its derivatives.	(Watson et al., 2008)
pAW8	Empty	pAW8	Cloning of *loxp-pol1-loxM3* and its derivatives.	(Watson et al., 2008)
pAW12	Empty	pAW12	Cloning of *loxp-pol1-loxM3* and its derivatives.	(Watson et al., 2008)
pAW41	Empty	pAW41	Cloning of *loxp-pol1-loxM3* and its derivatives.	(Watson et al., 2008)
pKL712	Empty	9His-5FLAG-pRS306	To amplify up the *9His-5FLAG* tag.	Kind gift from Dr K. Labib. The pRS306 vector was originally described by (Sikorski and Hieter, 1989)
pRA1	Codon-optimized (Geneart®), truncated *pol1* gene (*Sz. pombe*)	pTWO-E (Based on p17-ET with an N-terminal His-tag)	Express recombinant His6-Pol1.	This study. The pTWO-E commercially sourced from Novagen. Described in (Garces et al., 2011)
pRA2	Codon-opimized (Geneart®), truncated *swi7-1* gene (*Sz. pombe*)	pTWO-E (Based on p17-ET with an N-terminal His-tag)	Express recombinant His6-Swi7-1.	This study.
pRA4	None	pMCSG32	Express recombinant MBP-His6 (control).	pMCSG32 sourced from DNASU (plasmid repository). Described in (Stols et al., 2002)

pRA5	Codon-optimized (Geneart®), truncated *pol1* gene (*Sz. pombe*)	pMCSG32	Express recombinant MBP-Pol1-His6 (control). The *pol1* gene was codon-optimized.	This study.
pRA6	Codon-opimized (Geneart®), truncated *swi7-1* gene (*Sz. pombe*)	pMCSG32	Express recombinant MBP-Swi7-1-His6 (control). The *swi7-1* gene was codon-optimized.	This study.
pRA7	Full pol1 gene, amplified from genomic DNA from *Sz. pombe*.	TOPO-Blunt II	Quick access to *pol1* gene.	This study.
pRA8	Full pol1 gene, amplified from genomic DNA from *Sz. pombe*. The gene is flanked by *loxP* and *loxM3* sites.	pAW8	To introduce pol1 into the donor FMA8 base strain. Used to quickly mutagenize the *pol1* gene.	This study. The pAW8 was described in (Watson et al., 2008)
pRA9	*pol1G1116E*. The gene is flanked by *loxP* and *loxM3* sites.	pAW8	To introduce *pol1G1116E* into the donor FMA8 base strain.	This study.
pRA10	*pol1G1116Q*. The gene is flanked by *loxP* and *loxM3* sites.	pAW8	To introduce *pol1G1116Q* into the donor FMA8 base strain.	This study.
pRA11	*pol1G1116S*. The gene is flanked by *loxP* and *loxM3* sites.	pAW8	To introduce *pol1G1116S* into the donor FMA8 base strain.	This study.
pRA12	*pol1G1116D*. The gene is flanked by *loxP* and *loxM3* sites.	pAW8	To introduce *pol1G1116D* into the donor FMA8 base strain.	This study.
pRA22	*9HIS-5FLAG-pol1*	pAW8	To tag *pol1*.	This study.

pRA23	*9HIS-5FLAG-swi7-1*	pAW8	To tag *swi7-1*.	This study.
Plasmids used for the *S. cerevisiae* study				
pCS11	Empty.	pYM-N24	To amplify the GAL1-3HA and insert in pRS305 to create pCS25.	The pYM-N24 was first described in (Janke et al., 2004)
pCS14	*GAL1-TAP-Ø*	pRS305	To integrate *GAL1-TAP-Ø* in the genome at *LEU2*.	This study. The pRS305 vector was originally described by (Sikorski and Hieter, 1989)
pCS25	*GAL1-3HA-Ø*	pRS305	To integrate *GAL1-3HA-Ø* in the genome at *LEU2*.	This study.
pCS26	*SEN1 (2-931)*	pRS305-*GAL1-3HA* (pCS25)	To integrate *GAL1-3HA-SEN1 (2-931)* in the genome at *LEU2*.	This study.
pCS30	*SEN1 (2-931)*	pRS305-*GAL1-TAP* (pCS14)	To integrate *GAL1-TAP-SEN1 (2-931)* in the genome at *LEU2*.	This study.
pCS31	*SEN1 (2-1103)*	pRS305-*GAL1-TAP* (pCS14)	To integrate *GAL1-TAP-SEN1 (2-1103)* in the genome at *LEU2*.	This study.
pCS32	*SEN1 (931-2231)*	pRS305-*GAL1-TAP* (pCS14)	To integrate *GAL1-TAP-SEN1 (931-2231)* in the genome at *LEU2*.	This study.
pCS33	*SEN1 (1095-2231)*	pRS305-*GAL1-TAP* (pCS14)	To integrate *GAL1-TAP-SEN1 (1095-2231)* in the genome at *LEU2*.	This study.
pCS39	*SEN1 (2-2231)*	pRS305-*GAL1-TAP* (pCS14)	To integrate *GAL1-TAP-SEN1 (2-2231)* in the genome at *LEU2*.	This study.
pCS40	*SEN1 (2-1901)*	pRS305-*GAL1-TAP* (pCS14)	To integrate *GAL1-TAP-SEN1 (2-1901)* in the genome at *LEU2*.	This study.
pCS42	*SEN1 (2- 622)*	pRS305-*GAL1-3HA* (pCS25)	To integrate *GAL1-3HA-SEN1 (2- 622)* in the genome at *LEU2*.	This study.
pCS43	*SEN1 (410- 931)*	pRS305-*GAL1-3HA* (pCS25)	To integrate *GAL1-3HA-SEN1 (410- 931)* in the genome at *LEU2*.	This study.

pCS47	*SEN1 (2-2231)*	pRS305-*GAL1-3HA* (pCS25)	To integrate *GAL1-3HA-SEN1 (2-2231)* in the genome at *LEU2*.	This study.
pCS56	*SEN1 (410- 913)*	pRS305-*GAL1-3HA* (pCS25)	To integrate *GAL1-3HA-SEN1 (410- 913)* in the genome at *LEU2*.	This study.
pCS57	*SEN1 (410- 761)*	pRS305-*GAL1-3HA* (pCS25)	To integrate *GAL1-3HA-SEN1 (410- 761)* in the genome at *LEU2*.	This study.
pCS58	*SEN1 (410- 501)*	pRS305-*GAL1-3HA* (pCS25)	To integrate *GAL1-3HA-SEN1 (410- 501)* in the genome at *LEU2*.	This study.
pCS59	*SEN1 (501- 931)*	pRS305-*GAL1-3HA* (pCS25)	To integrate *GAL1-3HA-SEN1 (501- 931)* in the genome at *LEU2*.	This study.
pCS60	*SEN1 (761- 931)*	pRS305-*GAL1-3HA* (pCS25)	To integrate *GAL1-3HA-SEN1 (761- 931)* in the genome at *LEU2*.	This study.
pCS61	*SEN1 (622-931)*	pRS305-*GAL1-3HA* (pCS25)	To integrate *GAL1-3HA-SEN1 (622-931)* in the genome at *LEU2*.	This study.
pCS115	*SEN1* promoter	pCS25 (pRS305-derivative)	Switch *SEN1* promoter for *GAL1* promoter in pCS25. To integrate *SEN1-3HA-Ø* in the genome at *LEU2*.	This study.
pCS116	*ACT1* promoter	pCS25 (pRS305-derivative)	Switch *ACT1* promoter for *GAL1* promoter in pCS25. To integrate *ACT1-3HA-Ø* in the genome at *LEU2*.	This study.
pCS117	*SEN1 (931-2231)*	pRS305-*SEN1-3HA* (pCS115)	To integrate *SEN1-3HA-SEN1 (931-2231)* in the genome at *LEU2*.	This study.
pCS118	*SEN1 (931-2231)*	pRS305-*ACT1-3HA* (pCS116)	To integrate *ACT1-3HA-SEN1 (931-2231)* in the genome at *LEU2*.	This study.
pCS119	*SEN1 (2-2231)*	pRS305-*SEN1-3HA* (pCS115)	To integrate *SEN1-3HA-SEN1 (2-2231)* in the genome at *LEU2*.	This study.
pCS120	*SEN1 (2-2231)*	pRS305-*ACT1-3HA* (pCS116)	To integrate *ACT1-3HA-SEN1 (2-2231)* in the genome at *LEU2*.	This study.
pCS123	*SEN1 (2-2231) W773A E774A W777A*	pRS305-*ACT1-3HA*	To integrate *SEN1 (2-2231) W773A E774A W777A* in the genome at *LEU2*. Obtained by site-directed mutagenesis of pCS120.	This study.

pCS124	*SEN1 (2-2231) L656A S657A K658A I659A L660A*	pRS305-*ACT1-3HA*	To integrate *SEN1 (2-2231) L656A S657A K658A I659A L660A* in the genome at *LEU2*. Obtained by site-directed mutagenesis of pCS120.	This study.
pCS125	*SEN1 (2-2231) D850A E851G V852A L853G L854A*	pRS305-*ACT1-3HA*	To integrate *SEN1 (2-2231) D850A E851G V852A L853G L854A* in the genome at *LEU2*. Obtained by site-directed mutagenesis of pCS120.	This study.
pCS128	*SEN1 (2-2231) V746G D747G P748G I749G*	pRS305-*ACT1-3HA*	To integrate *SEN1 (2-2231) V746G D747G P748G I749G* in the genome at *LEU2*. Obtained by site-directed mutagenesis of pCS120.	This study.
pKL205	*TAP-Ø*	pRS306	To amplify the TAP tag to make a fusion PCR of GAL1-TAP. This PCR was then inserted in pRS305 to create pCS14.	Kind gift from Dr K. Labib.
pKL935	*GAL1-Ø*	pRS306	To amplify the *GAL1* promoter to make a fusion PCR of GAL1-TAP. This PCR was then inserted in pRS305 to create pCS14.	Kind gift from Dr K. Labib.

Table 2.5. List of oligoes used in this study.

Oligo Name	5'-3' Sequence	Purpose
Oligonucleotides used in the *Sz. pombe* study		
NRT1	AAAAAAAAACCCGGGGTGGTGGTGGTATGGATGGT AGCCTGTTTTTCTTC	Input XmaI site. For cloning *swi7* in pMCSG32.
NRT2	AAAAAAAAACTCGAGACCACCACCGCCTTGGAAAT AGAGATTTTCACCACCACCGCTGGTTTCATGGCTAT AGTATTT	Remove STOP codon. Input XhoI site. For cloning *swi7* in pMCSG32
NRT3	TTCATTATCTTTAACCGCCTGGAAAAAAACCCGGAA GATTATCCG	Mutate glycine residue into glutamic acid (for codon-optimized *swi7* gene). pTWO-E/ pMCSG32 GGC to GAA
NRT4	CGGATAATCTTCCGGGTTTTTTTCCAGGCGGTTAAA GATAATGAA	Mutate glycine residue into glutamic acid (for codon-optimized *swi7* gene). pTWO-E/ pMCSG32 GGC to GAA
NRT5	TTGGTTGTAGCAGTCCGCAGAATAT	Sequencing to ensure mutation of codon-optimised *swi7* gene to *swi7-1* gene in vector.
NRT6	TATAGCGTTCTGCTGAGTCGTCTGA	Sequencing to ensure mutation of codon-optimised *swi7* gene to *swi7-1* gene in vector.
NRT7	TGGTAGTCTGTGCCTGTTTGGTAAA	Sequencing to ensure mutation of codon-optimised *swi7* gene to *swi7-1* gene in vector.

NRT8	GGTGCAGAAGATGGTCTGCAAGAAG	Sequencing to ensure mutation of codon-optimised *swi7* gene to *swi7-1* gene in vector.
NRT9	CTGGCAGATCAGATGGGTCTGCAGG	Sequencing to ensure mutation of codon-optimised *swi7* gene to *swi7-1* gene in vector.
NRT10	GTTCGTGTGGGTGATGTTATTCCGT	Sequencing to ensure mutation of codon-optimised *swi7* gene to *swi7-1* gene in vector.
NRT11	TTAGCTGGTTTCATGGCTATAGTAT	Sequencing to ensure mutation of codon-optimised *swi7* gene to *swi7-1* gene in vector.
NRT12	CGCTGCTGCTGCTCGGACGAGGCAG	Sequencing to ensure mutation of codon-optimised *swi7* gene to *swi7-1* gene in vector.
NRT13	GTGCCGTAGCGCTGAAGTCTTACG	Sequencing to ensure *pol1* and *swi7-1* genes are in frame with first ATG from maltose binding protein.
NRT14	CTCGGGAGCAAAGATAGAGGCTACGCTAGGGACG GTCGAGGAGGCAACCATTTTCGACGAGATAGTATT CGTGCTTCAAGTATTTCCCGTTAGAATACTCAAGCT TGGAC	Integration of loxP site 162 bp upstream of *swi7*
NRT15	ACAAGTAATACTGACCCTAGCTCACCACCACAAGA GAGTTAAACCGGAATAGGTTAGGGTTGTAAAGACA GTCACAAGATGTAAACTAATTTCACCACCCCCGCC GCCCG	Integration of loxP site 162 bp upstream of *swi7*

NRT16	CTTGGACTAATATGCCAAGCTTAAACCCTACCCTAAATTGACTGCTTTTTTATTAGAGAATTAAATTGCAAGCACATTTAGAATGTTAACAAGCTTAGCTACAAATCCCA	Integration of loxM3 immediately downstream of *swi7*
NRT17	AACAAGCTCACCGGAACGAATTTGAGGCCCAACGTTGGTTGCCATTTTGATATCTTTCCTCCCTACAGATGCGTTTGAGACACTGCGCAAGAATTCGAGCTCGTTTAAAC	Integration of loxM3 immediately downstream of *swi7*
NRT18	AAAAAAAGAGCTCTGAGAAGAGACTCGGGAGCAAAGATAGAGGCTACGCTAGGGACGGTCGAGGAGGCAACCATTTTCGACGAGATAGTATTCGTGCTTCAAG	PCR-amplify *swi7* from genomic DNA and add a SacI cut site upstream
NRT19	AAAAAAAACTAGTGTTAACATTCTAAATGTGCTTGCAATTTAATTCTCTAATAAAAAAGCAGTCAATTTAGGGTAGGGTTTAAGCTTGGCATATTAGTCCAAG	PCR-amplify *swi7* from genomic DNA and add a SpeI cut site downstream
NRT20	CTCGGGAGCAAAGATAGAG	Sequence and colony-PCR of *Sz. pombe* after transformation with products from product of pAW41. Also used after transformation with products from pAW5 and pAW12.
NRT21	TGAAATTAAACGTGAGTA	Sequence and colony-PCR of *Sz. pombe* after transformation with products from product of pAW41.

NRT22	CCTTATGTGTGTGCTCTGG	Sequence and colony-PCR of *S.pombe* after transformation with pAW5. Also used after transformation pAW12-product.
NRT23	AACTTGGTTATAAACATTG	Sequence and colony-PCR of *Sz. pombe* after transformation with products from product of pAW12. Also usable with final product.
NRT24	AACAAGCTCACCGGAACGA	Sequence and colony-PCR of *Sz. pombe* after transformation with products from product of pAW12. Also usable with final product.
NRT25	TAGGGTCAGTATTACTTGT	Sequence and colony-PCR of S.pombe. Used along with CRELOX_seq_swi7_start (antisense)
NRT26	ACGTTTTAAGTAAACGGTTCT	Sequence and colony-PCR of mainly FMA8 but compatible with wildtype.
NRT27	GTCGAGGATTCTCGCACAACC	Sequence and colony-PCR of mainly FMA8 but compatible with wildtype.
NRT28	TTCCATTTGGATAATCTTCTGGATTTTTTTCCAAACG ATTGAATATTATAAATTTGTTA	Site-directed mutagenesis of Pol1. G1116E (LoxP-LoxM3 system)
NRT29	TAACAAATTTATAATATTCAATCGTTTGGAAAAAAAT CCAGAAGATTATCCAAATGGAA	Site-directed mutagenesis of Pol1. G1116E (LoxP-LoxM3 system)
NRT30	GTCTTTCCATTTGGATAATCTTCTGGATTTTTCTGCA AACGATTGAATATTATAAATTTGTTAGCCG	Site-directed mutagenesis of Pol1. G1116Q (LoxP-LoxM3 system)

NRT31	CGGCTAACAAATTTATAATATTCAATCGTTTGCAGAAAAATCCAGAAGATTATCCAAATGGAAAGAC	Site-directed mutagenesis of Pol1. G1116Q (LoxP-LoxM3 system)
NRT32	CTTTCCATTTGGATAATCTTCTGGATTTTTGCTCAAACGATTGAATATTATAAATTTGTTAGC	Site-directed mutagenesis of Pol1. G1116S (LoxP-LoxM3 system)
NRT33	GCTAACAAATTTATAATATTCAATCGTTTGAGCAAAAATCCAGAAGATTATCCAAATGGAAAG	Site-directed mutagenesis of Pol1. G1116S (LoxP-LoxM3 system)
NRT34	TCTTTCCATTTGGATAATCTTCTGGATTTTTATCCAAACGATTGAATATTATAAATTTGTTAGC	Site-directed mutagenesis of Pol1. G1116D (LoxP-LoxM3 system)
NRT35	GCTAACAAATTTATAATATTCAATCGTTTGGATAAAAATCCAGAAGATTATCCAAATGGAAAGA	Site-directed mutagenesis of Pol1. G1116D (LoxP-LoxM3 system)
NRT40	AGAAGAGAGAGTAGTTGAAG	Determine mating type (common oligo)
NRT41	ACGGTAGTCATCGGTCTTCC	Determine mating type (specific to *mat1P*)
NRT42	TACGTTCAGTAGACGTAGTG	Determine mating type (specific to *mat1M*)
MK01	TGACAGATCTCCTTCTAACTCTGATTCCGA	Sequencing of *pol1* gene in *vivo*.
MK02	TAGTTCACTGTGAAATTGACACATTCTTTA	Sequencing of *pol1* gene in *vivo*.
MK03	GTGAGAGGTCTGAAGTTTCGTTGCTTAATA	Sequencing of *pol1* gene in *vivo*.
MK04	CTAATACCGCCTTATTCGAGCAGTTTGTCT	Sequencing of *pol1* gene in *vivo*.
MK05	AAAAACATATGGATGGTTCTTTGTTTTTTTTTTGGATGGAT	Sequencing of *pol1* gene in *vivo*.
Mk06	AAAAGGATCCTCAAGAAGTCTCATGTGAGTAGTATTTTCGAGC	Sequencing of *pol1* gene in *vivo*.

MK07	CCATGCTCTTCGAATTGG	Sequencing of *pol1* gene in *vivo*.
MK08	ATTTCTTGCGGACTATAAGC	Sequencing of *pol1* gene in *vivo*.
CS440	GCATACATTATACGAAGTTATGCATGCGGAGAGCT CTACGCTAGGGACG	Tagging of Pol1 (fission yeast) at the amino-terminal with a 5X FLAG, 9X His tag.
CS441	GATAAGTTCACGAGAAGACTTTAAAGTCGAA	Tagging of Pol1 (fission yeast) at the amino-terminal with a 5X FLAG, 9X His tag.
CS442	TTCGACTTTAAAGTCTTCTCGTGAACTTATCATGGG TGCTCATCACCACCATCACCATCATCA	Tagging of Pol1 (fission yeast) at the amino-terminal with a 5X FLAG, 9X His tag.
CS443	CGGCTTTATCCCCGCGTTTCTCTTTCTGGCACCCG CTCCAGCGCCTGCACCAGCTCC	Tagging of Pol1 (fission yeast) at the amino-terminal with a 5X FLAG, 9X His tag.
CS444	AGAAAGAGAAACGCGGGGATAAAGCCGTCT	Tagging of Pol1 (fission yeast) at the amino-terminal with a 5X FLAG, 9X His tag.
CS445	ATTGGTTGACTGTCGACATTTTCTTCTGTCTCGAGT GCAG	Tagging of Pol1 (fission yeast) at the amino-terminal with a 5X FLAG, 9X His tag.
CS784	CGTATAGCATACATTATACG	Aligns with loxP locus (either on pRA8 or pRA8-derived plasmids or when integrated into the genome). Used to amplify pol1 (Sz. pombe) from progeny of the FMA12 strain.
CS785	CGTATATAATACCATATACG	Aligns with loxM3 locus (reverse complemented, either on pRA8 or pRA8-derived plasmids or when integrated into the genome). Used to

		amplify pol1 (Sz. pombe) from progeny of the FMA12 strain.
CS786	ATCGAAGAACGTTCTAGCTG	Aligns to within the *pol1* gene (4705-4724 in pRA8 and pRA8-derived plasmids).
CS787	GCCGTACATCTCGGTATAATCC	Aligns to within the *pol1* gene (5407-5428 (RC) in pRA8 and pRA8-derived plasmids).
CS788	CCTCAACCTGCTAGCAGTCTTCC	Aligns within the *pol1* gene (5009-5031 in pRA8 and pRA8-derived plasmids).
CS789	ACAAATTGGCTACTTCGTCG	Aligns within the *pol1* gene (5590-5609 in pRA8 and pRA8-derived plasmids).
Oligonucleotides used in the *S. cerevisiae* study		
CS290	TGTATGGCATCTATCTCTATATATATAAAAAAGCGC ATCTGTTTATTATAACGTACGCTGCAGGTCGAC	Tagging/gene deletion of the *SEN1* gene in W303 (*S. cerevisiae*).
CS291	TATACACCAATATATATGCAGGTATAATTCCTAACA CTTTTACTTCAAGATCAATCGATGAATTCGAGCTCG	Tagging/gene deletion of the *SEN1* gene in W303 (*S. cerevisiae*).
CS292	CGGAATGCTTCATCTAGCCCATTTATCCCAAAAAAA AGAAAGCCTAGATCACGTACGCTGCAGGTCGAC	Tagging/gene deletion of the *SEN1* gene in W303 (*S. cerevisiae*).
CS293	ATTATTATTATTAATGTTGTTGCTATTATTATTATCAG GATTGTTGGAATTCATCGATGAATTCTCTGTCG	Tagging/gene deletion of the *SEN1* gene in W303 (*S. cerevisiae*).
CS294	CTTCGCATATTTTAGGATCTTG	Checking N- and C-terminal tagging of the *SEN1* gene in W303 strains.

CS295	CATTTATCACAAACAGAGAG	Checking N- and C-terminal tagging of the *SEN1* gene in W303 strains.
CS296	TCAAATAGTGTTTTATCCGG	Checking N- and C-terminal tagging of the *SEN1* gene in W303 strains.
CS297	AAACAGCCCTCGACTCCCTC	Checking N- and C-terminal tagging of the *SEN1* gene in W303 strains.
CS297	AAACAGCCCTCGACTCCCTC	Checking N- and C-terminal tagging of the *SEN1* gene in W303 strains.
CS355	TTACAGCGGCCGCGAGCTCTAGTACGGATTAG	Cloning *GAL1-3HA* in pRS305 vector.
CS356	AGATAAGTCGACACCGGCACCGGCACCAGC	Cloning *GAL1-3HA* in pRS305 vector.
CS361	TTACTGTCGACAATTCCAACAATCCTGATAAT	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS362	CAATGGCTCGAGTCACTTTTCTTTTCTGTAGTTTTCAG	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS363	CAATGGCTCGAGTCATTCTAACTCCTGTTTAGCCAAT	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.

CS364	CAATGGCTCGAGTCATACCTCATCGGGGCCTTG	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS365	CAATGGCTCGAGTCATGATCTAGGCTTTCTTTTTTTTG	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS366	TTACTGTCGACAAGGAATTATCAAGACTTGGG	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS367	TTACTGTCGACGCTGAATTGGCTAAACAGGAG	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS368	ATGAATTCCAACAATCCTG	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS369	TAACATTTTGCACTTCCCT	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS370	GACATCGTATGCCAATGGA	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.

CS371	TCGAGATTACTACCGGTTT	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS372	ATGGCGTTCCTGCGTATAA	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS373	CATTGCTGGCCAAACAGCT	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS374	TCTGGTGTGTTTGCTAATT	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS375	CAGGCGAGTGAGATTCTAC	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS376	GACGATAGCAGTAGTGAGG	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS377	GTCGCTAAACAAGTCATAC	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.

CS378	TTTCTTTAATTCAGGGTCC	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS379	GGTAATCCAGAAAGTCCAA	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS380	CTGGATGTTCAATACCGTA	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS381	ACGAAATCTAGTGTCGGTT	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS382	CAGCACCTCATCTGGTACA	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS383	AGGTACCTGCTGCTATTAC	Checking *SEN1* constructs under *GAL1-3HA/ GAL1-TAP* in pRS305-derived vector by sequencing.
CS446	GTTACTCTAGAATGAAAGCTGATGCGCAAC	Expressing TAP-tagged *SEN1* fragment under control of a *GAL1* promoter.

CS447	CGAGGTCGACGGCACCCGCTC	Expressing TAP-tagged *SEN1* fragment under control of a *GAL1* promoter.
CS448	AATATCCTCGAGGTCACTTTTCTTTTCTGTAGTTTTCAG	Expressing fragments of *SEN1*.
CS449	AATATCCTCGAGGTCATTCTAACTCCTGTTTAGCCAAT	Expressing fragments of *SEN1*.
CS450	AATATCCTCGAGGTCATACCTCATCGGGGCCTTG	Expressing fragments of *SEN1*.
CS451	AATATCCTCGAGGTCATGATCTAGGCTTTCTTTTTTTTG	Expressing fragments of *SEN1*.
CS581	TTATTATTAATGTTGTTGCTATTATTATTATCAGGATTGTTGGAATTGGCACCCGCTCCAGCGCCTG	To construct a *SEN1* Degron
CS679	GTGAGGAGAGTGACAACGATATAG	To make a full length construct of Sen1 under the control of the *GAL1* promoter and tagged at the N-terminal with 3HA.
CS680	AAAGGGAACAAAAGCTGGGTAC	To make a full length construct of Sen1 under the control of the *GAL1* promoter and tagged at the N-terminal with 3HA.
CS687	TACACCTACAAAAAAGCTCTACTTTGTCATTTATTTTCTACTTCATCTCTATTAAGGCGCGCCAGATCTG	To construct a heat-inducible degron of *SEN1*. To be used with oligo CS 688. This oligo pair can also be used for other N-terminal modification of SEN1.

CS688	TTATTATTAATGTTGTTGCTATTATTATTATCAGGATTGTTGGAATTCATGGCACCCGCTCCAGCGCCTG	To construct a heat-inducible degron of *SEN1*. To be used with oligo CS 687. This oligo pair can also be used for other N-terminal modification of SEN1.
CS712	CAATGGCTCGAGTCACTCAACAATTGGATTGTCTT	Cloning of SalI_*SEN1 (2-410)*_XhoI.
CS713	CAATGGCTCGAGTCATGACGGTACTATCAAAAGAC	Cloning of SalI_*SEN1 (2-622)*_XhoI.
CS714	TTACTGTCGACGAGCATCAAACAGAAGTTTAC	Cloning of SalI_*SEN1(410-931)*_XhoI
CS747	AGCTGGTGCAGGCGCTGGAGCGGGTGCCGTCGACGAGCATCAAACAGAAGTTTAC	Forward primer. Used to clone fragments of SEN1 starting with residue 410 in pCS14. Gibson compatible. Also includes the SalI restriction site.
CS748	AGCTGGTGCAGGCGCTGGAGCGGGTGCCGTCGACACAGTACTGTTGACTAAAACAG	Forward primer. Used to clone fragments of *SEN1* starting with residue 501 in pCS14. Gibson compatible. Also includes the SalI restriction site.
CS749	AGCTGGTGCAGGCGCTGGAGCGGGTGCCGTCGACAAGAGCCAAAATACCGAAAAGG	Forward primer. Used to clone fragments of *SEN1* starting with residue 761 in pCS14. Gibson compatible. Also includes the SalI restriction site.
CS750	CTAAAGGGAACAAAAGCTGGGTACCGGGCCCCTACTTTTCTTTTCTGTAGTTTTCAG	Reverse primer. Used to clone fragments of *SEN1* ending with residue 931 in pCS14. Gibson compatible. Also includes the PspXI restriction site.

CS751	CTAAAGGGAACAAAAGCTGGGTACCGGGCCCCTAGAATATTTTAGCCTTTTGTAGAAT	Reverse primer. Used to clone fragments of *SEN1* ending with residue 913 in pCS14. Gibson compatible. Also includes the PspXI restriction site.
CS752	CTAAAGGGAACAAAAGCTGGGTACCGGGCCCCTACAAAGTTTGGAAATTAGCAAACA	Reverse primer. Used to clone fragments of *SEN1* ending with residue 761 in pCS14. Gibson compatible. Also includes the PspXI restriction site.
CS753	CTAAAGGGAACAAAAGCTGGGTACCGGGCCCCTATTCAAATTCATCATTCTTATAAAGC	Reverse primer. Used to clone fragments of *SEN1* ending with residue 501 in pCS14. Gibson compatible. Also includes the PspXI restriction site.
CS754	AGCTGGTGCAGGCGCTGGAGCGGGTGCCGTCGACTCAAATACTGCCGTAGCTGAG	Forward primer. Used to clone fragments of *SEN1* starting with residue 622 in pCS14. Gibson compatible. Also includes the SalI restriction site.
CS755	GAGATTCATTCCCAATCAAACTAATCAAAATTCAAGACAATCCAATTGTTATTAAGGCGCGCCAGATCTG	To produce a PCR (with oligo CS756) from plasmid pKL506. This PCR is transformed into a diploid and causes a disruption mutation of *SEN1*.
CS756	GCCGGCACAGAATCCGTTAAGTCGATCACCTTCCCAAGTCTTGATAATTCATCGATGAATTCGAGCTCG	To produce a PCR (with oligo CS755) from plasmid pKL506. This PCR is transformed into a

		diploid and causes a disruption mutation of *SEN1*.
CS757	GCAGTTTAGCATTAGCTGATG	To check (by sequencing) whether the PCR generated from oligoes CS755 and CS756 engendered the correct disruption in the *SEN1* gene in one of the chromatids (in a diploid). Oligo CS371 should also be used.
CS758	GGATCTGGTGCTGGTTCAGGAGCAGGTTCAGGTGA ATTATCAAGACTTGGGAAGGTG	To produce a PCR product from gDNA (obtained from strain #1353 or other *SEN1-TAP* strains). This oligo aligns with bases 2794- 2814 in the *SEN1* ORF (thus beginning with residue 931). To be used in combination with oligo CS291. The PCR will be used for a substrate of another PCR reaction.
CS759	GAGATTCATTCCCAATCAAACTAATCAAAATTCAAG ACAATCCAATTGTTGGATCTGGTGCTGGTTCAGG	Using a PCR generated from gDNA coming from strain # 1353 (with oligoes CS758 and CS291), this oligo in tandem with oligo CS291 will be used to produce another PCR that aligns with bases 1178-1228 in SEN1 ORF. It includes a linker, followed by another region that aligns with the *SEN1* gene. This PCR will be used to loop out

		the *URA3* gene that had been used to disrupt the *SEN1* gene previously, creating a gene that has lost bases that translate to residues 410-931.
CS760	CACGTAATAGTCAAACTCTTTCACG	To check (by sequencing, along with other oligoes) whether the bases encoding residues 410- 931 have been successfully removed from SEN1. Designed by Giacomo.
CS761	CCGTTAACGATAATACGGG	To sequence *SEN1*.
CS762	CCGCATATGTTGAGGCTT	To sequence *SEN1*.
CS769	GCGGATTACAGAGACACTCG	To sequence *SEN1*.
CS771	CAATCGACGGTTTCCAAGGTC	To check *SEN1* (1-1901) tagging.
CS819	AAACAGCTGCTAGTATCTTTAAAAAATATTAATGGT CTTTTGATAGTACCGGAATTATCAAGACTTGGGAAG	To make *SEN1 (622-931Δ)* with no linker.
CS820	GGATCTGGTGCTGGTTCAGGAGCAGGTTCAGGTGA ATTATCAAGACTTGGGAAGG	To make *SEN1 (622-931Δ)* with a short linker.
CS821	AAACAGCTGCTAGTATCTTTAAAAAATATTAATGGT CTTTTGATAGTACCGGGATCTGGTGCTGGTTCAG	To make *SEN1 (622-931Δ)* with a short linker.
CS822	TGGATCTGGTGCAGGAGGCTCTGGTGAATTATCAA GACTTGGGAAG	To make *SEN1 (622-931Δ)* with a long linker.
CS823	GGATCTGGTGCTGGTTCAGGAAGTGGAGGCGCTG GGGCAGGATCTGGATCTGGTGCAGGAGGCTC	To make *SEN1 (622-931Δ)* with a long linker.

CS824	AACAGCTGCTAGTATCTTTAAAAAATATTAATGGTCTTTTGATAGTACCGGGTGGAAGCGGTGGCAGCATCGACAACTTCGAC	To make *SEN1 (622-931Δ)* with a CIP linker.
CS825	AGCATCGACAACTTCGACGACATCCTGGGCGAGTTCGAGGGTGCTGGATCTGGTGAATTATCAAGACTTGGGAAG	To make *SEN1 (622-931Δ)* with a CIP linker.
CS826	AGCATCGACAACTTCGACGACATCCTGGGCGAGTTCGAGGGTGCTGGATCTGGTAAAGCACTCACGGAAGAGG	To make *SEN1 (622-913Δ)* with a CIP linker.
CS827	AACAGCTGCTAGTATCTTTAAAAAATATTAATGGTCTTTTGATAGTACCGAAAGCACTCACGGAAGAGG	To make *SEN1 (622-913Δ)* with no linker.
CS828	GGATCTGGTGCTGGTTCAGGAGCAGGTTCAGGTAAAGCACTCACGGAAGAGG	To make *SEN1 (622-913Δ)* with a short linker.
CS829	TGGATCTGGTGCAGGAGGCTCTGGTAAAGCACTCACGGAAGAGG	To make *SEN1 (622-913Δ)* with a long linker.
CS830	GAGATTCATTCCCAATCAAACTAATCAAAATTCAAGACAATCCAATTGTTGGTGGAAGCGGTGGCAGCATCGACAACTTCGACG	To make *SEN1 (410-931Δ)* with a CIP linker.
CS831	GAGATTCATTCCCAATCAAACTAATCAAAATTCAAGACAATCCAATTGTTAAAGCACTCACGGAAGAGG	To make *SEN1 (410-913Δ)* with no linker.

CS832	GAGATTCATTCCCAATCAAACTAATCAAAATTCAAG ACAATCCAATTGTTGGATCTGGTGCTGGTTCAG	To make *SEN1 (410-931Δ)* with a short linker.
CS833	GAGATTCATTCCCAATCAAACTAATCAAAATTCAAG ACAATCCAATTGTTGAATTATCAAGACTTGGGAAG	To make *SEN1 (410-931Δ)* with no linker.
CS898	GTATGGCATCTATCTCTATATATATAAAAAAGCGCA TCTGTTTATTATAAATTAAGGCGCGCCAGATCTG	To make *SEN1* starting from 912 or 930
CS899	GACTCACTTAATTTTATCATTAAAAGTACAAAAAATA TCGATGAATTCGAGCTCG	To make *SEN1* starting from 912 or 930
CS900	ATTTTTTGTACTTTTAATGATAAAATTAAGTG	To make *SEN1* starting from 912 or 930
CS901	TCTTTTCTGTAGTTTTCAGCTTCTGTAGCGACCTCT TCCGTGAGTGCTTTATTCATTTATAATAAACAGATG CGCTTTTT	To make *SEN1* starting from 912 or 930
CS902	GGCACAGAATCCGTTAAGTCGATCACCTTCCCAAG TCTTGATAATTCCTTCATTTATAATAAACAGATGCGC TTTTTT	To make *SEN1* starting from 912 or 930
CS929	ACGGCAGTATTTGACGGTACTATCCCACCACCACC ACCATTTTTTAAAGATACTAGCAGCTGTTTGGCCAG CAATGGACCGT	SEN1_ SIte directed mutagenesis of LSKIL (614-618) to AAAAA
CS930	ACGGTCCATTGCTGGCCAAACAGCTGCTAGTATCT TTAAAAAATGGTGGTGGTGGTGGGATAGTACCGTC AAATACTGCCGT	SEN1_ SIte directed mutagenesis of LSKIL (614-618) to AAAAA

CS933	GTAAACTTATAAATCGTATCCAAGAATAACGCACAT GATGCCGCGAATTTTAAAAATTCCTTTTCGGTATTTT GGCTC	SEN1_ SIte directed mutagenesis of WESCW (773-777) to AASCA
CS934	GAGCCAAAATACCGAAAAGGAATTTTTAAAATTCGC GGCATCATGTGCGTTATTCTTGGATACGATTTATAA GTTTAC	SEN1_ SIte directed mutagenesis of WESCW (773-777) to AASCA
CS935	CAGACGTGCTTATAATTAATCTAACACAAGATTCCG CACCTGCTCCGGCGCTCAATCTCAGCCAATACAAC ATATTCTTA AAG	SEN1_ SIte directed mutagenesis of DEVLL (850-854) to AGAGA
CS936	CTTTAAGAATATGTTGTATTGGCTGAGATTGAGCGC CGGAGCAGGTGCGGAATCTTGTGTTAGATTAATTAT AAGCACGT CTG	SEN1_ SIte directed mutagenesis of DEVLL (850-854) to AGAGA
CS939	GTACTTGGCCATCATTTCTCCCAAGGAACCACCCA CTTTCACATGCTTTTC	SEN1_ SIte directed mutagenesis of DDSLV (876-880) to GGSLGEM
CS940	GAAAAGCATGTGAAAGTGGGTGGTTCCTTGGGAGA AATGATGGCCAAGTAC	SEN1_ SIte directed mutagenesis of DDSLV (876-880) to GGSLGEM
CS941	AAGTTTGGAAATTAGCAAACACACCAGAACCCCCG CCTCCAAATGCACTCACTACGTCCATCAAAACA	SEN1_ SIte directed mutagenesis of FVDPISGV (746-753) to FGGGGSGV(746-753)

CS942	TGTTTTGATGGACGTAGTGAGTGCATTTGGAGGCG GGGGTTCTGGTGTGTTTGCTAATTTCCAAACTT	SEN1_ SIte directed mutagenesis of FVDPISGV (746-753) to FGGGGSGV(746-753)
CS953	ATCTAAGCGGCCGCACAAGCGCGCCTCTACCT	Cloning *ACT1* promoter instead of *GAL1.*
CS954	TACAATACTAGTTGTTAATTCAGTAAATTTTCGAT	Cloning *ACT1* promoter instead of *GAL1.*
CS955	ATCTAAGCGGCCGCCGTTATGTGACCAATGTATA	Cloning *SEN1* promoter instead of *GAL1.*
CS956	TACAATACTAGTTTATAATAAACAGATGCGCTTT	Cloning *SEN1* promoter instead of *GAL1.*
CS981	AGAAGATGATTATAAGCTACCC	Clone TAP by Gibson in plasmid.
CS982	GGAACAAAAGCTGGGTACCGAGGCCTTATACACCA ATATATATGCAGG	Clone TAP by Gibson in plasmid.

2.4 BIOCHEMISTRY

2.4.1 GENOMIC DNA EXTRACTION

Individual clones were grown overnight in 5 ml of either YPD or YEA medium. Cells were washed in sterile d.H_2O and re-suspended in 200 µl of lysis buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 3% (w/v) SDS) to which 200 µl of sterile d.H_2O, 200 µl of phenol/chloroform/alcohol (25:24:1) and a small quantity of glass beads were added. The cells were lysed by vigorous shaking for 1 min. The mixture was then centrifuged at 12,000 *g* for 2 min. 380 µl of the clear supernatant was transferred to a tube containing 760 µl of 100% ethanol. The resulting solution is around 70% (v/v) ethanol and the DNA precipitates. The DNA is recovered by centrifugation and is washed in fresh 70% (v/v) ethanol to remove any residual phenol/chloroform contaminant that may inhibit downstream applications and recovered again by centrifugation. The supernatant was discarded. The pellet was dried at 37°C for 30 min before resuspension in 50 µl of TE supplemented with 50 µg/ml RNAse A. The mixture was left to incubate at 37°C for at least 1 h before storage at −20°C.

2.4.2 TCA PROTEIN EXTRACTION

Individual clones were grown overnight in 5 ml of either YPD or YEA medium. Cells were washed in sterile d.H_2O and re-suspended in 300 µl of 20% (w/v) TCA to which a small quantity of glass beads was added. The cells were then lysed by vigorous shaking for 1 min. The supernatant was then transferred to a new Eppendorf and left-over protein precipitates were fished out by washing the cell debris and glass beads with 300 µl 5% (w/v) TCA and this solution was then added to the 30% (w/v) TCA collected previously. The 600 µl combined solution was centrifuged for 10 min at 3000 *g*. The supernatant was discarded and the protein pellet was re-suspended in 100-200 µl of 1 x Laemmli buffer supplemented with 150 mM Tris base. The sample was boiled for 4 min before centrifugation at 3000 *g* for 10 min. The supernatant was then kept at −20°C

for long-term storage and analysed using electrophoresis on polyacrylamide gels.

2.4.4 IMMUNO-PRECIPITATION OF TAP-TAGGED PROTEINS FROM YEAST SAMPLES

Dilute or concentrated yeast samples were prepared as described in section 2.1.7. For dilute samples, frozen cell pellets were lysed using a mechanized pestle and mortar at −80°C (Spex Sample Prep, 6870). The resulting powdered yeast cells were thawed and kept at 4°C. 1 g of the lysate was assumed to be equivalent to 1 ml. To each volume of thawed lysate was added one quarter of volume plus 50 µl of a solution of 50% (v/v) glycerol, 100 mM Hepes-KOH (pH 7.9), 50 mM potassium acetate, 50 mM magnesium acetate, 0.5% Igepal® CA-630 (Sigma), 2mM EDTA supplemented with 2 mM glycerophosphate, 2 mM sodium fluoride, 1 mM DTT, 1% (v/v) Sigma protease inhibitor cocktail and 0.24% (w/v) EDTA-free Complete Protease Inhibitor Cocktail (Roche). Pierce Universal Nuclease was added to a final concentration of 0.4 U/µl. The mixture was transferred to a centrifuge tube and left on a rotating platform for 30 min to allow degradation of nucleic species. The lysate was then clarified from cell debris by step-wise centrifugation first at 18,700 *g* and then at 126,600 *g* in a vacuum. The supernatant was isolated from the cell debris and 50 µl of this cell extract was added to 100 µl of 1.5 X Laemmli buffer, boiled for 4 min and flash-frozen on dry-ice.

The remainder of the cell extract was incubated with 100 µl of TAP beads (M-270 Dynabeads® Epoxy beads (Invitrogen) bound to rabbit anti-sheep IgG (S1265, Sigma)) in two separate Eppendorfs. The mixture was placed on a rotating platform at 4°C for 2 h before washing the beads with solutions of 100 mM Hepes-KOH (pH 7.9), 50 mM potassium acetate, 50 mM magnesium acetate, 2 mM EDTA, 0.1% (v/v) Igepal® CA-630 thrice. Importantly, the first wash was supplemented with 2 mM glycerophosphate, 2 mM sodium fluoride, 1 mM DTT, 1% (v/v) Sigma protease inhibitor cocktail and 0.24% (w/v) EDTA-free Complete Protease Inhibitor Cocktail (Roche). After the three washes, 50

µl of 1 x Laemmli buffer was added to each 50 µl of beads, homogenized by vigorous vortexing and boiled for 4 min. 50 µl from each duplicate was added to one another in order to average the samples, split again in volumes of 50 µl in two Eppendorfs and flash-frozen on dry-ice. The cell extracts and IP samples were then kept at −80°C for long-term storage.

For concentrated samples, frozen cell pellets were also lysed using the mechanized pestle and mortar at −80°C and the powdered lysate was allowed to thaw at 4°C. Again 1 g of the lysate was deemed to be equivalent to 1 ml and to each volume of thawed lysate was added a quarter of volume plus 50 µl of a solution of 50% (v/v) glycerol, 100 mM Hepes-KOH (pH 7.9), 50 mM potassium acetate, 50 mM magnesium acetate, 0.5% Igepal® CA-630 (Sigma), 2mM EDTA supplemented with 2 mM glycerophosphate, 2 mM sodium fluoride, 1 mM DTT, 1% (v/v) Sigma protease inhibitor cocktail and 0.24% (w/v) EDTA-free Complete Protease Inhibitor Cocktail (Roche). In addition, 1 ml of a solution of 10% (v/v) glycerol, 50 mM potassium acetate, 50 mM magnesium acetate, 0.1% (v/v) Igepal® CA-630, 1 mM DTT, 2mM DTT supplemented with 2 mM glycerophosphate, 2 mM sodium fluoride, 1 mM DTT, 1% (v/v) Sigma protease inhibitor cocktail and 0.24% (w/v) EDTA-free Complete Protease Inhibitor Cocktail (Roche). Pierce Universal Nuclease was added to a final concentration of 1.6 U/µl. The mixture was transferred to a centrifuge tube and left on a rotating platform for 40 min to allow degradation of nucleic acid species. The samples were then clarified by centrifugation and incubated with TAP beads, similar to dilute samples.

2.4.5 RECOMBINANT EXPRESSION OF *SZ. POMBE* POL1 IN *ESCHERICHIA COLI* BL21(DE3)

Individual clones of *E. coli* BL21 (DE3) carrying either the pRA4 (pMCSG32), pRA5 or pRA6 were inoculated in LB (lysogeny broth) medium consisting of 10% (w/v) tryptone, 5% (w/v) yeast extract and 10% (w/v) NaCl. Ampicillin to a final concentration of 100 µg/ml was used to prevent contamination and to prevent plasmid loss. To trigger expression of the recombinant genes, IPTG

was added to logarithmically-growing cultures to a final concentration of 1 mM. Upon addition of IPTG, the cultures were grown at 20°C for 6 h to favour transcription, translation and folding of the recombination proteins. Thereafter, cells were harvested by centrifugation at 4,000 × *g* for 20 min at 4°C. The cells were lysed by addition of lysozyme (to a final concentration of 1.5 mg/ml) and sonication. EDTA-free protease inhibitor tablets (Roche) and 10 mM PMSF (Fischer) were used to reduce digestion of the recombinant proteins during cell lysis. The resulting crude extract were centrifuged at 10,000 × *g* for 25 min at 4°C. The supernatant (cleared lysate) contained the recombinant proteins.

The cleared lysates were then diluted 1:6 in column buffer (20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 1 mM EDTA, pH 8, 1mM DTT) supplemented with 1 tablet of EDTA-free Complete Protease Inhibitor Cocktail (Roche) and 10 mM PMSF. The diluted samples were then mixed with three volumes of amylose resin (NEB) pre-washed with 5 column volumes of column buffer. The diluted lysate and resin were incubated at 4°C for 2 h on a rotary shaker. After incubation, the mixture was packed in an empty gravity flow column (Econo-Pac® Chromatography Columns, Biorad) and the column buffer was allowed to separate from the resin by gravity flow at 4°C, ensuring that the resin did not dry. The resin was then washed with 12 columns volumes of column buffer by gravity flow and the flow-through was collected for analysis on acrylamide gels. The MBP-tagged proteins were then eluted by using 10 successive 1/5th column volumes fractions of elution buffer supplemented (20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 1 mM DTT, 10 mM maltose). A small quantity of each elution was saved for analysis on acrylamide gels.

The first three fractions were pooled and diluted 1:10 in 50 mM Hepes buffer with no additional salt (pH 7). The diluted samples were then fed manually through a syringe into separate 1 ml pre-packed HiTrap™ Heparin HP columns (GE Healthcare) pre-washed with 10 column volumes of 50 mM Hepes buffer (pH 7). Once the samples were applied, the columns were washed with 5 column volumes of 50 mM Hepes buffer (pH 7). The flow-through was collected for analysis on acrylamide gels. The proteins were eluted by washing successively with 3 column volumes of 50 mM Hepes buffer

(pH 7) with 0.5 M NaCl, 3 column volumes of 50 Hepes buffer (pH 7) with 1 M NaCl, 3 columns volumes of 50 mM Hepes buffer (pH 7) with 1.5 M NaCl and 3 columns volumes of 50 mM Hepes buffer (pH 7) with 2 M NaCl. The different elutions were pooled according to the concentration of NaCl used and small quantities of the pooled samples were analysed on an acrylamide gel. The samples eluted with 0.5 and 1 M NaCl solutions contained the highest quantity of fusion proteins. The first two elutions for either concentrations were pooled together (4 ml of material) and diluted 1:10 in a solution of 50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole and 0.2% (v/v) IGEPAL CA-630.

2 ml of of 50% Ni-NTA slurry was then washed twice with one column volume of 50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole and 0.2% (v/v) IGEPAL CA-630 to remove the preservatives. The slurry was resuspended in a solution of 50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole and 0.2% (v/v) IGEPAL CA-630 to a final volume of 2 ml. The elutions from the heparin column were then added to the slurry and allowed to incubate on a rocking platform at 4°C for 2 h. Thereafter, the mixture was packed in an empty gravity flow column (Econo-Pac® Chromatography Columns, Biorad) and the column buffer was allowed to separate from the resin by gravity flow at 4°C, ensuring that the resin did not dry. The flow-through was collected for analysis on acrylamide gels. The column was washed with 8 column volumes of 50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole and 0.2% (v/v) IGEPAL CA-630. The recombinant proteins were eluted using 4 column volumes of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole and 0.2% (v/v) IGEPAL CA-630). The 4 column volumes were collected separately. The first two elutions contained the highest levels of the recombinant proteins, for both the wildtype and mutant proteins. The recombinant proteins were kept in 50% (v/v) glycerol at −20°C and in 25% (v/v) glycerol at −70°C for long-term storage.

2.4.6 DETECTION OF PROTEINS BY WESTERN BLOTTING, COOMASSIE BLUE OR SILVER STAINING

Protein were separated according to size by running protein samples through 5, 6, 7, 8 or 10% denaturing polyacrylamide gels (National Diagnostics) in electrophoretic reactions. For IP samples used for mass spectrometric analysis, pre-cast Novex™ Wedgewell™ 4-12% Tris-Glycine polyacrylamide gels were used to separate proteins to minimize human contamination of the samples. To simply assess the presence or relative quantities of proteins, the proteins were stained either with Coomassie or silver. For Coomassie staining, gels were rinsed in deionized water before staining with InstantBlue Protein Stain (Expedeon) at 25°C for at least 1 h whilst, for silver staining, gels were fixed for 10 min in a solution of 40% (v/v) methanol and 13.5% (v/v) formaldehyde before rehydration in deionized water. The gels were then sensitized in a solution of 0.02% (w/v) sodium thiosulphate for 1 min before briefly rinsing in deionized water. Finally, the gels were then incubated in a solution of 0.1% (w/v) $Ag(NO_3)$ for 10 min, rinsed with deionized water before development in a solution of 3% (w/v) sodium carbonate, 0.05% formaldehyde and 0.0004% (w/v) sodium thiosulphate. Upon achieving satisfactory signal to noise ratios, the reaction was stopped by adding an excess of citric acid to the solution.

In order to quickly identify or relatively quantify specific proteins within a sample, proteins were transferred to either a PVDF or a nitrocellulose membrane using a semi-dry method after separation on polyacrylamide gels. The membranes were blocked in 5% (w/v) skimmed-milk at 25°C for 1 h to reduce unspecific binding. Thereafter, the membranes were incubated either at 4°C overnight or at 25°C for 1 h in 5% (w/v) milk supplemented with the appropriate primary antibody (Table 2.6). The membranes were then washed thrice in TBST for 5 min. If the primary antibody was conjugated to the HRP (horseradish peroxidase) enzyme, the membranes were then treated with the Western blotting Reagents (ECL). Otherwise, the membranes were treated with 5% milk supplemented with a secondary antibody conjugated to the HRP

enzyme. The membranes were then washed thrice in TBST for 5 min and treated with the Western blotting Reagents (ECL). The chemiluminescent signal was detected by exposing chemiluminescent films (Hyperfilm^{TM}, ECL) to them. The secondary antibodies used were all commercially sourced. These are anti-sheep IgG-HRP (Sigma, A3415), anti-rabbit IgG-HRP (Rockland, 18-8816-33) and anti-mouse IgG-HRP (Cell Signaling Technology, #7076). Unless otherwise stated, all primary antibodies recognize *S. cerevisiae* proteins specifically.

Table 2.6. Primary antibodies used in this study. * The antibody used against MBP was used as a primary antibody. After probing with this antibody, membranes were washed and incubated with secondary anti-mouse antibody.

Antibody	Typical concentrations	Host	Provenance
anti-Pol2	1:1,000	Sheep	Kind gift from K. Labib.
anti-Pol1	1:1,000	Sheep	Kind gift from K. Labib.
anti-Mcm6	1:20,000	Sheep	Kind gift from K. Labib.
anti-Mcm5	1:1,000	Sheep	Kind gift from K. Labib.
anti-Mcm4	1:1,000	Sheep	Kind gift from K. Labib.
anti-Mcm3	1:1,000	Sheep	Kind gift from K. Labib.
anti-Ctf4	1:20,000	Sheep	Kind gift from K. Labib.
anti-Pob3	1:1,000	Sheep	Kind gift from K. Labib.
anti-Cdc45	1:2,500	Sheep	Kind gift from K. Labib.
anti-Dpb2	1:1,000	Sheep	Kind gift from K. Labib.
anti-Csm3	1:1,000	Sheep	Kind gift from K. Labib.
anti-Sld5	1:500	Sheep	Kind gift from K. Labib.
anti-Psf1	1:500	Sheep	Kind gift from K. Labib.
anti-TAP-HRP (PAP)	1:1,000,000- 1:25,000	Rabbit	Sigma (P1291)
anti-c-MYC (9E10)	1:5,000- 1:500	Mouse	Sigma (M4439)
anti-FLAG (M2)	1:5,000- 1:500	Mouse	Sigma (F3165)

anti-HA (12CA5)	1:5,000- 1:500	Mouse	Sigma (11 583 816 001)
anti-MBP-HRP*	1:30,000	Mouse	Commercially available from NEB (E8038). Kind gift from M. Balasubramanian.

2.4.7 FACS ANALYSIS BY FLOW CYTOMETRY

Yeast cells at a density of 0.7 x 10^7 cells/ml were fixed in 70% (w/v) ethanol. The suspension could be kept at 4°C for long-term storage. To process the samples, 200 µl of suspension was added to 1 ml of fresh 50 mM sodium citrate solution and vortexed vigorously. The cells were pelleted by centrifugation at 3000 *g* for 3 min, re-suspended in 1 ml of fresh sodium citrate and vortexed vigorously. This removes any alcohol and allows the cells to rehydrate. To remove any RNA molecules, the cells were then pelleted and re-suspended in fresh 50 mM sodium citrate supplemented with 0.1 mg/ml RNase A. The suspension was then allowed to incubated at 37°C overnight. The cells were again recovered by pelleting and re-suspended in 500 µl of pepsin in 50 mM HCl. The suspension was centrifuged and the pellet was re-suspended in 1 ml of 50 mM sodium citrate supplemented with 2 µg/ml propidium iodide. Before processing using a FACScan (BD), the cells were tip-sonicated to prevent clumping and vortexed to promote homogeneity within the suspension. Excitation of fluorophores is carried with a 488 nm laser light and the emission photons were detected using a 650 nm long pass filter.

2.4.8 PRIMER EXTENSION ASSAY

Polymerase activity from purified protein samples was measured using a hairpin DNA/RNA hybrid oligonucleotide. The DNA/RNA hybrid oligo mimics the *in vivo* substrate of Polymerase α where RNA acts as the primer and DNA as the template. The oligonucleotide was labelled at the 5' end with radioactive ^{32}P using T4 Polynucleotide kinase (PNK) and ^{32}P- γ-ATP. Reaction mixtures (20 µl) contained 50 mM Hepes medium (adjusted to pH 7.0), 1.0 mM

magnesium acetate, 75 mM KCl, 50 mM sodium glutamate, 0.05% (v/v) IGEPAL CA-630, 1 mM DTT, 0.1 mg/ml BSA, 20 μM dNTPs and 10-40% (v/v) extract of interest to assess the latter's polymerase activity. The reactions were allowed to run for 0.5- 256 min at either 25°C, 30°C or 37°C before termination with 20 μl of formamide loading buffer. The samples were heated to 92°C to denature the DNA and were run on 8 or 10% polyacrylamide sequencing gels (7 M urea). The radioactive fragments separated on the gel were detected using films. Exposure times varied from overnight to seven days at −70°C.

2.4.9 Electrophoretic motility shift assay (EMSA)

Standard electrophoretic mobility shift assay reactions were conducted with the 5' radioactively-labelled primer used in Section 2.4.8. In brief, equal amounts of nucleic acid substrate were incubated with different quantities of different protein samples in a 20 μl reaction of 10 mM Tris-HCl (pH 7.5), 25 mM NaCl, 2.5% (v/v) glycerol, 0.5 mM EDTA and 2 mM DTT. Non-specific competitor DNA (sonicated salmon sperm) was added to a final concentration of 0.1 mg/ml. The protein samples were allowed to bind to the substrate for 30 min at 4°C. 3 μl of sucrose dye solution (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF and 40% (w/v) sucrose) was added to each 20 μl reaction. The reaction was then separated on 6% native polyacrylamide gel. The gel was pre-run at 100 V for 30 min and subsequently run at 100 V after sample loading in 0.5× TBE buffer (89 mMTris-acetate, 89 mM boric acid, 2 mM EDTA) at 4°C until the dye end was close to the end of the gel. The gel was then dried and the radioactive fragments separated on the gel were detected using films. Exposure times varied from overnight to seven days at −70°C.

2.4.9 Determining protein concentrations using the bicinchoninic acid assay

The bicinchoninic acid (BCA) protein assay is a two-component and highly precise colorimetric assay used to quantify proteins. Unlike the Bradford assay, it is compatible with most ionic and non-ionic detergents and other contaminants. The first step involves the reduction of copper ions (Cu^{2+}) to cuprous ions (Cu^{+}) in an alkaline environment by protein side-chains (cysteine, tyrosine and tryptophan). However, contrary to the Bradford reagent, the peptide backbone also contributes to the reduction of copper ions. This minimizes the variability across samples with different protein compositions, making the BCA assay superior to the Bradford assay. The contribution of the peptide backbone increases at higher temperatures (37°C and 60°C), increasing the sensitivity of the assay. The second step involves the chelation of cuprous ions by BCA (two BCA molecules chelate one cuprous ion), leading to colour formation. A strong purple product is formed and the assay exhibits a strong linear absorbance at 562 nm with increasing protein concentrations. The concentration of recombinant protein samples was measured using a commercially-sourced Pierce™ BCA Protein Assay Kit (ThermoFischer), using serial dilutions of bovine serum albumin (BSA) as standards.

2.5 MASS SPECTROMETRIC ANALYSIS

2.5.1 IMMUNO-PRECIPITATION AND MASS SPECTROMETRIC METHODS

To obtain sufficiently concentrated protein samples for mass spectrometric analysis, four independent harvest of 7 x 10^9 cells carrying the *GAL1-TAP-SEN1 (2-931)* allele and grown in YPGAL were conducted for each phase of the cell cycle (G_1, S and G_2). For the control (no bait), four independent harvests of 2 x 10^{10} cycling cells, encoding the *GAL1-TAP-∅* allele, were carried out. The cycles of the cell population at each different harvest were verified by FACS. The samples were lysed by using a motorized pestle and mortar cooled to −80°C using liquid nitrogen. The lysate was thawed to 4°C and clarified by centrifugation at 12,500 rpm in a high speed centrifuge and at 32,500 rpm in an ultra-centrifuge. TAP-Sen1 (2-931) was immuno-precipitated using IgG bound to Dynabeads in a 100 mM Hepes, 50 mM potassium acetate solution. After three washes, TAP-Sen1 (2-931) and its co-precipitates were cleaved off the Dynabeads using AcTEV protease at 24°C for 2 h. After cleavage, the resulting CBP-Sen1 (2-931) and its co-precipitates were incubated with pre-washed calmodulin beads at 4°C for 2 h and, after washing, the proteins were removed from the calmodulin resin by boiling in 30 μl of 1 x Laemmli buffer. This simultaneously concentrates the proteins. The samples from the four biological replicates were then pooled together. Thereafter, the samples were run for a small distance on a commercially-sourced 4-12% acrylamide gel to reduce the potential for contamination especially from keratin. The gel was then cut in thin slices and processed using a robot (MS Bioworks, USA). The protein samples were successively washed, reduced, alkylated and digested by trypsin. The reaction was then analysed using a nanoLC/MS/MS (Waters NanoAcquity HPLC/ThermoFischer Q Exactive). Of the ions identified, the fifteen most abundant were selected for tandem MS. To process the data, the spectra obtained were compared to the spectra of *in silico* trypsinized *Saccharomyces cerevisiae* proteins (SGD database).

2.5.2 ANALYSIS OF RAW DATA

The raw data obtained from the mass spectrometric screen is presented in Table S1 (Supplementary Data). The raw data (number of peptides) for selected proteins is presented in graphical form in Figures 4.10, 4.13 and 4.15. For Figure 4.11, the data were processed as indicated below.

Firstly, it was assumed that the mass spectrometry apparatus was not saturated with ions. Then, for both the bait and every protein identified in the screen, the molecular weight (M), length of primary sequence of the protein (L) and the number of tryptic sites + 1 (T) was determined using the SGD database and the ExPASy PeptideCutter software (http://web.expasy.org/peptide_cutter/). For each protein, the following variable (X) was calculated: [X=M*(T/L)]. The number of specific peptides (N) was calculated by subtracting the number of non-specific peptides in the control sample from the total number of peptides in each of the test samples. The number of specific peptides (N) was then divided by the X variable. The resulting variable (Y) gives an output value normalized for both the molecular weight of the protein and the number of tryptic sites dispersed along its length. Finally, the value of the Y variable was normalized to that of Sen1 (2-931) ($Y/Y^{Sen1\ (2\text{-}931)}$) in that specific test sample. For instance, the Y value of $Pol1_{G1}$ was divided by the Y value of Sen1 $(2\text{-}931)_{G1}$.

2.6 IMAGE ACQUISITION AND PROCESSING

2.6.1 IMAGE ACQUISITION OF FILMS

Films were exposed to chemiluminescent or radioactive signals for the appropriate length of time. The films were then developed and scanned using with an Epson V700 Scanner. The image was acquired in 8-bit grayscale at a resolution of 300 dpi (dots/pixels per inch).

2.6.2 IMAGE ACQUISITION OF COLONIES GROWING ON PLATES

Cells were grown on the appropriate medium for the required number of days. When required, plates were placed upright on an Epson V700 Scanner to be scanned. The image was acquired in 8-bit grayscale at a resolution of 600 dpi (dots/pixels per inch).

2.6.3 IMAGE PROCESSING

Acquired images were opened with Adobe Photoshop CS6 image software. The images were adjusted for brightness and contrast, then rotated and cropped as required. The images were processed to remove dust and scratches and despeckled to 1 pixel. Finally, the images were formatted to widths of 3 or 6 cm (without altering the width to length ratio) and the resolution was set to 508 dpi.

2.6.4 IMAGE GENERATION

Data were analysed with Microsoft Excel and R Studio (*ver* 1.0.153). Graphical figures were created using R Studio. Images were created in Adobe Illustrator CS6.

CHAPTER 3: INVESTIGATING THE ROLE OF POLYMERASE α IN THE IMPRINTING PROCESS AT THE MPS1 LOCUS IN THE MODEL ORGANISM SCHIZOSACCHAROMYCES POMBE.

3.1 BACKGROUND

The imprinting process in *Schizosaccharomyces pombe* is essential for cell-type switching and is dependent on a number of *cis*-regulatory elements and *trans*-acting (protein) factors. Several of these *trans*-acting factors were isolated by screening for mutants with reduced frequencies for cell-type switching (Gutz and Schmidt, 1985). Others were identified at later dates, using a multitude of genetic screens (Codlin and Dalgaard, 2003, Holmes et al., 2012, Zech et al., 2015).

Among the genes required for the imprinting process, components of the fork protection complex (*swi1*^{TOF1}, *swi3*^{CSM3} and *mrc1*) are required for both fork stalling at the *MPS1* locus and for fork arrest at the *RTS1* locus. In addition, *swi7* is necessary for imprinting at *MPS1* in an ill-characterized but pausing-independent manner (Dalgaard and Klar, 2000, Dalgaard and Klar, 2001, Singh and Klar, 1993, Zech et al., 2015). Importantly, to date, only one switching-defective mutant of the *swi7* gene (*swi7-1*) has been isolated, suggesting that this defect is specific to this one mutant and not to some general disruption of the gene. Work in the Klar lab have since shown that *swi7-1* is allelic to the *pol1* gene and carries a missense mutation (G3347A) (Singh and Klar, 1993). The resulting protein (Pol1G1116E) is thought to have reduced affinity for its DNA substrate by virtue of the increased negative charge at the mutated residue (Perera et al., 2013). How this translates to reduced frequency of cell-type switching is uncertain and represents a gap in scholarship.

Here, I have used *in vitro* assays to characterize the defects of *swi7-1*. The Swi7-1 protein was found to have reduced affinity for a DNA/RNA hairpin substrate that mimics the natural substrate of Pol1. This reduced affinity also severely affected the polymerase activity of the protein. These observations provide some clues about the role of Polymerase α at the *MPS1* locus during the imprinting process. I have also identified an allele of $spp1^{PRI1}$ that is characterized by reduced cell-type switching. Spp1 is the catalytic primase subunit of Polymerase α, suggesting that the catalytic activity of the primase subunit of Polymerase α may be required for the imprinting process at the *MPS1* locus. The nature of the imprint at *MPS1* itself is still disputed. It is either a nick (Kaykov and Arcangioli, 2004) or two ribonucleotides (Vengrova and Dalgaard, 2004, Vengrova and Dalgaard, 2006). Involvement of primase in the imprinting process could be interpreted as evidence of the RNA-nature of the imprint.

3.2 *SWI7-1* IS UNIQUELY SWITCHING-DEFECTIVE AMONGST *POL1* MUTANTS

To date, only the *swi7-1* allele of *pol1* has been shown to be defective for cell-type switching, suggesting a very specific defect. To test the specificity of this phenotype, I have crossed strains carrying different mutants of *pol1* with the h^{90}+ :: *LEU2* marker (referred henceforth to as h^{90LEU2}), including *swi7-1*, *pol1-1* (D'Urso et al., 1995), *pol1-ts13* (Bhaumik and Wang, 1998) and *pol1-H4* (Murakami and Okayama, 1995) (Table 3.1). Here, the 1.1 kb *S. cerevisiae LEU2* marker replaces a stretch of 1.5 kb DNA distal (with *cen2* as a reference point) to the *mat1* locus so that the distance between the *mat1* and *mat2P* loci is not substantially altered and that the imprinting at *MPS1* is unaffected (Fig 3.1). This not only ensures that autotrophy for leucine is a marker for homothallism (the ability to switch cell-types) but also that recombination between the receptor *mat1* locus and either of its donor loci (*mat2P* and *mat3M*) is not compromised (Zech et al., 2015)

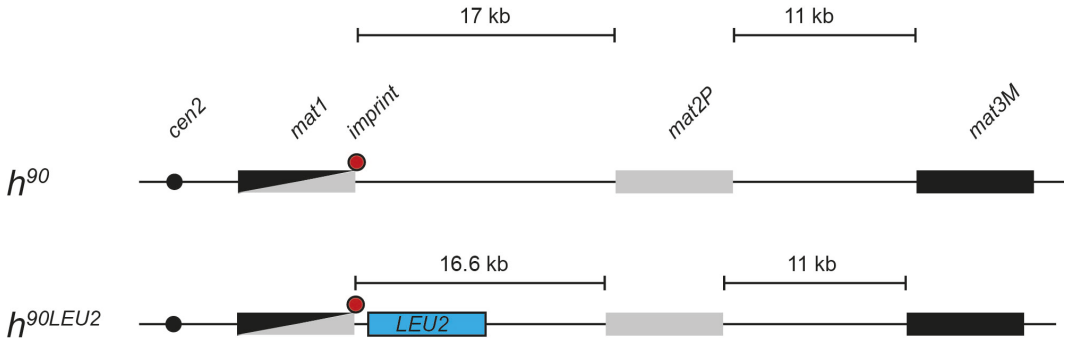


Figure 3.1. The h^{90LEU2} marker allows simple selection of homothallism by genetically linking it to leucine autotrophy. This marker has been described in Zech *et al.* (2015).

Table 3.1. Summary of *pol1* alleles tested for defects in cell-type switching.

Allele	Residues Mutated	Brief Overview	Phenotype	Reference
pol1-1	Not characterized.	Temperature-sensitive allele with an elongated cell (cdc) phenotype at non-permissive temperatures and following HU treatment. A small proportion of the cell population undergoes abnormal mitosis at non-permissive temperatures. The allele confers a mutator phenotype.	Disrupted catalytic activity. Temperature-sensitive. Sensitive to hydroxyurea. Mutator phenotype.	(D'Urso et al., 1995)
pol1-ts13	470-472Δ	Temperature-sensitive allele that leads to arrest in early to mid S phase at non-permissive temperatures. The allele also confers a mutator phenotype.	Disrupted catalytic activity. Temperature-sensitive. Cut phenotype. Decreased silencing of mating donor loci. Mostly insensitive to hydroxyurea. Mutator phenotype	(Bhaumik and Wang, 1998) (Ahmed et al., 2001)

pol1-H4	G889D	Temperature-sensitive allele with an elongated cell (cdc) phenotype. Cells cannot undergo DNA replication at non-permissive temperatures. Has a similar phenotype to mcl1-101/cos1 mutants with reduced propagation of centromeric DNA. This allele confers a mutator phenotype.	Disrupted catalytic activity Temperature-sensitive Cut phenotype. Decreased silencing of mating donor loci. Increased duration of mitotic S phase. Mutator phenotype	(Murakami and Okayama, 1995) (Ahmed et al., 2001)
swi7-1	G1116E	Viable mutant. Leads to impaired response to alkylation damage (such as MMS) in S phase. Confers a mutator phenotype.	Cut phenotype. Decreased silencing of mating donor loci. Hypersensitive to hydroxyurea and MMS. Mutator phenotype.	(Singh and Klar, 1993)

The cell-type switching efficiency of each mutant was tested by growth on sporulation medium at 25°C (permissive temperature), followed by staining with iodine vapour. In homothallic (self-fertile or switching) strains, individual cells of complementary cell-types will mate, producing a transient diploid that quickly undergoes sporulation. The walls of these spores contain α-glucans and β-glucans starch molecules (Manners and Meyer, 1977) that can be stained with iodine vapours (Forsburg and Rhind, 2006). On the other hand, heterothallic (non-self-fertile or non-switching) strains exist as a homogeneous population cell-type-wise and these cells cannot mate with one another. Exposure to iodine vapour leads to light staining, producing uniformly yellow/white colonies. Switching-defective strains adopt an intermediate phenotype where cells switch cell-type inefficiently, allowing for localized mating and sporulation. Upon iodine staining, this leads to a characteristic speckled (or mottled) phenotype.

As shown previously (Zech et al., 2015), h^{90LEU2} cells have a phenotype similar to wildtype h^{90} cells. Cells carrying the *swi7-1* allele stained with the characteristic speckled (or mottled) phenotype and cells carrying both h^{90LEU2} and *swi7-1* also presented the same phenotype. Thus, the h^{90LEU2} construct does not interfere with the cell-type switching process. When present in cells carrying this h^{90LEU2} marker, the other mutant alleles of *pol1* led to dark staining after exposure to iodine vapour, indicating efficient cell-type switching (Fig 3.2). This suggests that *swi7-1* is uniquely switching-defective amongst *pol1* mutants.

However, cells carrying the *pol1-1* allele were characterized by some minor defects in cell-type switching. The extent of this defect was much lower compared to that of the *swi7-1* allele. The *pol1-1* mutant was initially isolated in a screen for temperature-sensitive mutations of *pol1* with checkpoint defects (D'Urso et al., 1995). However, at non-permissive temperatures, only a subpopulation of cells underwent aberrant mitoses and because of this low penetrance, the allele was deemed unsuitable. This low penetrance could also underpin the iodine staining pattern in h^{90LEU2} *pol1-1* cells where limited defects in Pol1-1 stochastically result in some cells switching inefficiently whilst, in

most cells, this defect is muted. This contrasts to *swi7-1* colonies where colonies stain with a homogeneous speckled phenotype. This suggests that the switching-defective phenotype of *swi7-1* arise from a specific defect and not some generalized loss of function of Pol1.

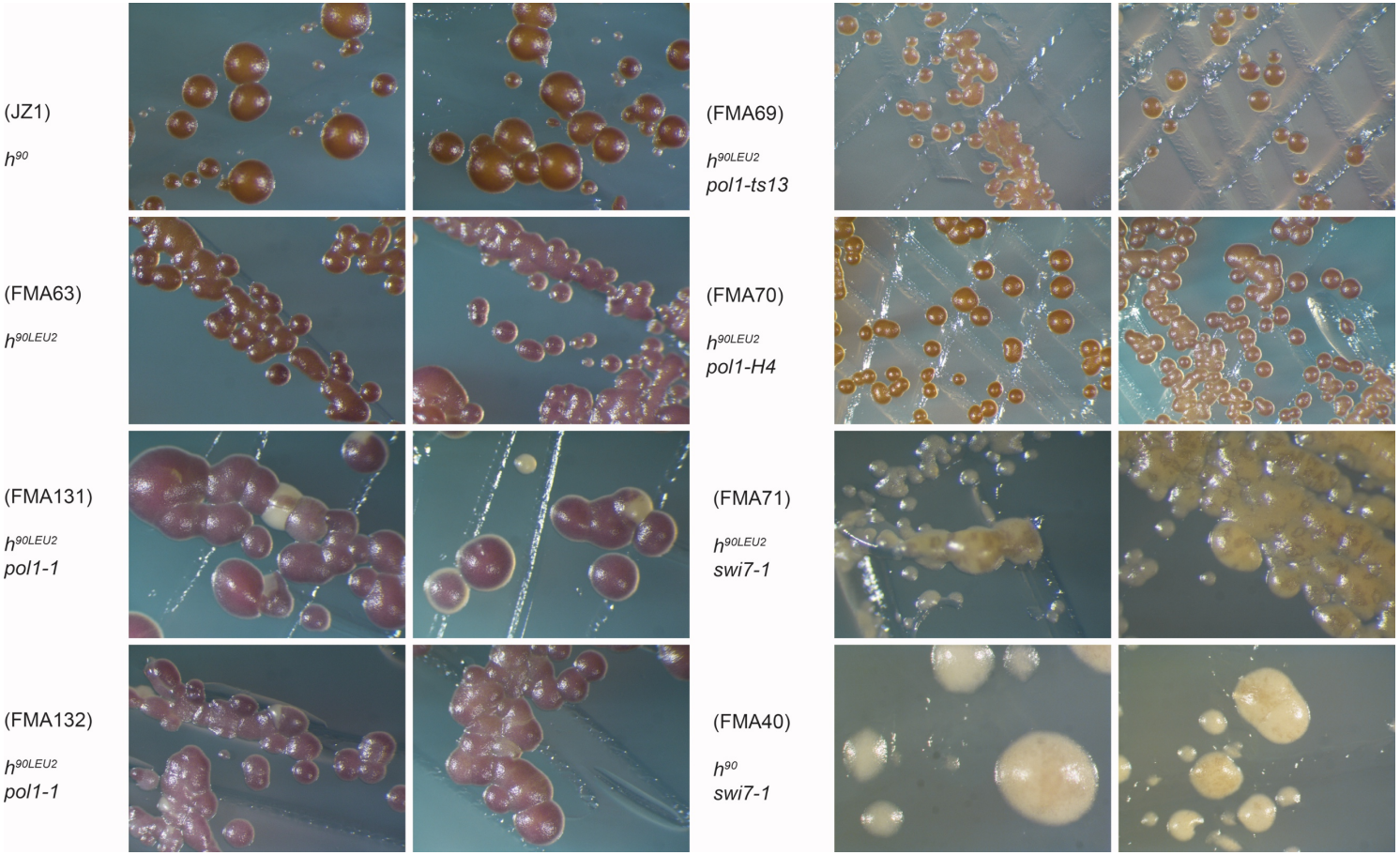


Figure 3 2. Iodine staining reveals that the *swi7-1* allele is uniquely switching-defective. However, the *pol1-1* allele has some mild switching-defects as well.

However, as the staining pattern of the *pol1-1* colonies could reflect unstable proteins, the switching- and imprinting-defects in *swi7-1* cells could be a result of differences in protein levels or in protein stability with wildtype cells. Given that the relative levels of Pol1 and Swi7-1 have not been assayed previously (Singh and Klar, 1993), we constructed strains carrying the *9His-5FLAG-pol1* and *9His-5FLAG-swi7-1* alleles using a recombinase mediated cassette technique (Watson et al., 2008). These strains were grown in non-selective medium and were then treated with cycloheximide to inhibit translation of new protein molecules. Prior to cycloheximide treatment, there was no difference in the levels of Pol1 and Swi7-1 (Fig 3.3). The levels of the two proteins were also indistinguishable up to 4 h after addition of cycloheximide to the medium. This suggests that Swi7-1 is expressed at similar levels to Pol1 and that the Swi7-1 protein is no more unstable than its wildtype counterpart.

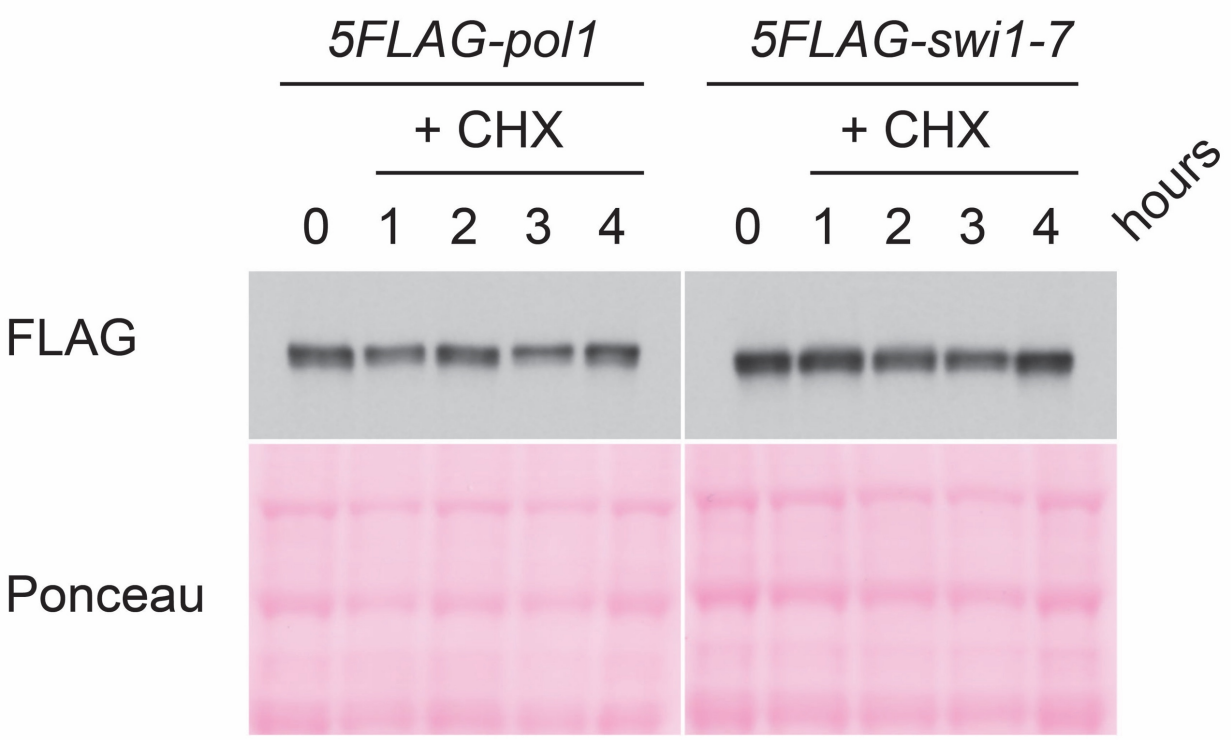


Figure 3.3. Strains carrying either the wildtype *pol1* or the *swi7-1* allele express similar levels of the protein and the mutant variant is no more unstable than Pol1. Strains of *Sz. pombe* were grown to a density of 0.7 x 10^7 cells/ml in YEA medium and treated with cycloheximide to a final concentration of 100 μg/ml up to 4 h. Samples were collected every hour for TCA analysis and run on a 6% acrylamide gel. Strains used: FMA**85** and FMA**86**.

3.3 CHARACTERIZING THE DIFFERENCES BETWEEN *POL1* AND *SWI7-1 IN VITRO*

Having established that the decreased cell-type switching in strains carrying the *swi7-1* allele was not reflective of decreased protein stability, we then undertook a reductive approach to characterize the differences between *pol1* and *swi7-1*. We first expressed truncated versions (G976-C3684, that encodes for residues 326 to 1228) of *pol1* and *swi7-1* tagged at the N-terminus with a hexa-histidine tag in the bacterial strain BL21 (DE3). Truncated proteins were used as the full-length enzymes are susceptible to degradation in the absence of their co-enzymes (Spb70, Spp1 and Spp2) (Perera et al., 2013). However, insufficiently high yields of proteins were obtained using this approach, perhaps as a result of mis-folding (not shown). We optimized expression of the proteins by codon-optimizing the genes for expression in bacteria and by using the *E. coli* maltose binding protein tag fused to the N-terminus of the constructs and a hexa-histidine tag at their C-termini. The MBP is known to enhance the solubility of its fusion partners either spontaneously or by acting as a magnet for chaperones (Raran-Kurussi and Waugh, 2012).

The resulting yields of the recombination proteins were markedly higher. These were purified from cleared bacterial lysate by successive passage through amylose, heparin and Ni-NTA columns. Samples from each purification steps were saved for analysis using Coomassie stain and Western blotting (Fig 3.4).

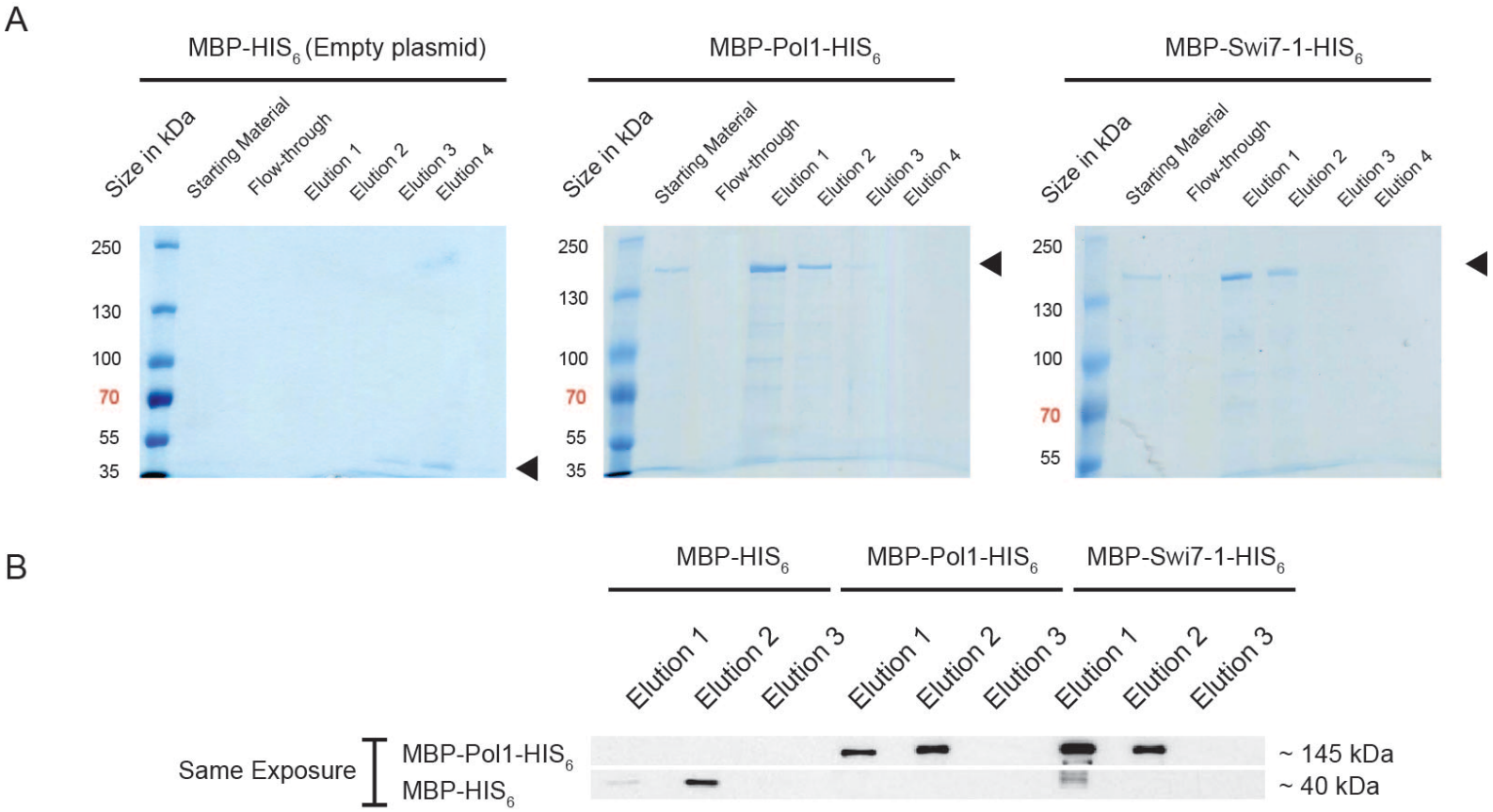

A
MBP-HIS6 (Empty plasmid)
MBP-Pol1-HIS6
MBP-Swi7-1-HIS6
Size in kDa
Starting Material
Flow-through
Elution 1
Elution 2
Elution 3
Elution 4
250
130
100
70
55
35
B
MBP-HIS6
MBP-Pol1-HIS6
MBP-Swi7-1-HIS6
Elution 1
Elution 2
Elution 3
Same Exposure
MBP-Pol1-HIS6
MBP-HIS6
~ 145 kDa
~ 40 kDa

Figure 3.4. Similar amounts of recombinant Pol1 and Swi7-1 proteins were purified for use in *in vitro* assays. Cleared bacterial lysate from BL21 (DE3) *E. coli* strains carrying either an empty vector (pRA4), a vector with the codon-optimized *pol1* gene (pRA5) or the codon-optimized *swi7-1* allele (pRA6) were purified by passage through three successive columns (amylose, heparin and Ni-NTA). Here, the samples for the last column purification (Ni-NTA) is shown. (**A**) Coomassie stain of the recombinant proteins. The arrows indicate the presence of either MBP-HIS_6, MBP-Pol1-HIS_6 or MBP-Swi7-1-HIS_6. Comparable amounts of recombinant Pol1 and Swi7-1 were recovered. (**B**) Western blot against MBP for the first three elutions obtained from the Ni-NTA column purification. The signals shown are from identical exposure times.

We then used the samples recovered after Ni-NTA purification to extend a DNA/RNA hairpin that mimics the natural substrate of Pol1 at 37°C (for increased enzymatic activity) for 1 h. The samples purified from bacteria carrying the empty control were unable to extend the substrate, nor were they able to degrade the nucleic species suggesting that the purification strategy used was stringent enough to remove any contaminating polymerases, helicases and nucleases of bacterial origin. Samples purified from bacteria carrying the truncated *pol1* gene were able to fully extend the hairpin DNA, beyond the extent achieved by the Klenow Polymerase (NEB). By contrast, no detectable polymerization of the DNA/RNA hairpin was observed when substituting the Swi7-1 protein for its wildtype counterpart (Fig 3.5). In order to better characterize the difference in activity of the mutant compared to the wildtype, different concentrations of the two proteins were used to extend the same DNA/RNA substrate at 37°C. Only at the highest concentration was discernible polymerizing activity observed with the Swi7-1 protein. Comparison between the enzymatic activities of the Pol1 and Swi7-1 proteins also suggests that Swi7-1 has around 2.5-5% polymerase activity compared to the wildtype enzyme (Fig 3.6).

We then used the primer extension assay at the highest concentration assayed for each enzyme to carry time-course experiments at 37°C. Although the *swi7-1* allele does not confer temperature-sensitivity, the time-course experiment was also carried at more physiological temperatures (24°C and 30°C) to ensure that the stability of the recombinant proteins were not compromised. These assays revealed that the Swi7-1 protein is much less functional than its wildtype counterpart at all three temperatures tested (Fig 3.7). The mutated residue does not correspond to the active site of the protein and a catalytic dead variant of the protein (Pol1D894N) renders cells inviable (Bhaumik and Wang, 1998). It is unlikely then that the decreased functionality associated with Swi7-1 is based on a reduction in catalytic activity. Instead, based on crystallographic data, the phenotype observed in Swi7-1 can be attributed to a predicted decreased affinity of the protein for its substrate (Perera et al., 2013). To test this hypothesis, the relative affinities of the

recombinant Pol1 and Swi7-1 for their substrates were compared, using the same DNA/RNA substrate used previously (in the primer extension assays) in an electrophoretic mobility shift assay (EMSA). In brief, the recombinant Pol1 and Swi7-1 proteins were allowed to interact with the primer substrate at 4°C and the sample reactions were immediately run on native gels at 4°C. The substrate on its own, or in the presence of a non-interacting protein, will run a certain distance, based on its size. However, binding to the substrate causes an upward shift as the resulting protein-DNA complex is heavier and migrates at a slower rate through the native gels.

Using the Klenow fragment, we first tested whether the primer substrate was suitable for use in a mobility shift assays. We found that, contrary to BSA, the Klenow fragment was able to bind to the substrate, causing a mobility shift (Fig 3.8A). In fact, under the conditions used, 1.9 pmol of Klenow (in a 20 µl reaction mix) was sufficient to cause an upward shift. The recombinant Pol1 construct also caused a shift, observable from 4.7 pmol of protein used. Unlike results from the Klenow fragment, two distinctive shifted bands were observable (Fig 3.8B). These bands may correspond to sequential loading of Pol1 molecules on the substrate, mutually exclusive occupancy state of the substrate (for example a substrate could either accommodate two protein molecules of Pol1 at either extreme of its DNA/RNA stretch, but only one molecule of Pol1 should the latter bind centrally to a permissible DNA/RNA region) or formation of protein-protein complexes. This contrasts to the Klenow fragment. Whilst titrating the Klenow down leads to a gradual loss of nucleic acid-protein complexes (reflecting an excess of substrate), no intermediate shifts were observable. The difference between the two proteins reflect the difference in substrate preference of the two proteins. Indeed, whilst the Klenow fragment binds to single-stranded DNA immediately past a primer (Turner et al., 2003), Pol1 binds almost exclusively to DNA/RNA hybrid substrates (Perera et al., 2013). Interestingly, we did not detect a shift indicating nucleic acid-protein interaction when substituting Swi7-1 for Pol1. This indicates that the mutant protein has reduced affinity for its substrate *in vitro* (Fig 3.8B, C).

The mutated residue in Swi7-1 aligns to the Ser1134 residue in budding yeast's Pol1 (Perera et al., 2013). Crystallographic data places this residue in the thumb domain of Pol1 and this domain is important for interaction with the RNA component of its DNA/RNA substrate. In line with the crystallographic data, the *in vitro* results presented here suggests that the G1116E mutation in Swi7-1 leads to decreased affinity for its substrate. Given the striking differences in activity between Pol1 and the mutant Swi7-1, it is remarkable that *swi7-1* cells are viable and show few signs of sickness unless when challenged by DNA damaging agent. This contradiction may suggest that, *in vivo*, the limitations of the Swi7-1 protein are compensated by the other subunits of Pol α and perhaps other components of the replisome.

Like Pol1, the Klenow fragment adopts a right-handed conformation where the analogous thumb domain is required for interaction with its substrates. Removal of 24 residues from the tip of the Klenow fragment's thumb domain triggers a 100-fold decrease in its affinity for its templates, as well as a 70-fold increase in the number of +1 frameshifts (Minnick et al., 1996). It is noteworthy then, that the *swi7-1* allele confers a mutator phenotype in *Sz. pombe* (Koulintchenko et al., 2012). It should also be noted that the Pol1-1, Pol1-H4 and Pol1-ts13 proteins have never been tested for either their ability to extend a DNA/RNA hybrid *in vitro* or for their ability to bind to such a substrate *in vitro*. Thus, there is a paucity of data to which to compare the results generated using the Pol1 and Swi7 proteins. However, given that the mutation in Pol1-ts13 is located in a different domain to that of the mutation found in Swi7-1 (Fig S.1), it is unlikely that Pol1-ts13 has any defects in binding to DNA/RNA hybrids or to extend them. By contrast, the Pol1-H4 has a mutation in the polymerase domain of Pol1, similar to Swi7-1 (Fig S.1). However, the mutated residue in Pol1-H4 (G889E) aligns to Gly904 in the *S. cerevisiae* homolog. This residue is not involved in binding of the protein to its substrates (Perera et al., 2013), suggesting that Pol1-H4 will behave differently from Swi7-1 although this would need to be investigated empirically. Finally, the mutation in Pol1-1 has yet to be characterized. Should the mutation occur within the polymerase domain, it is possible that Pol1-1 may have some defects in substrate binding and in polymerase activity.

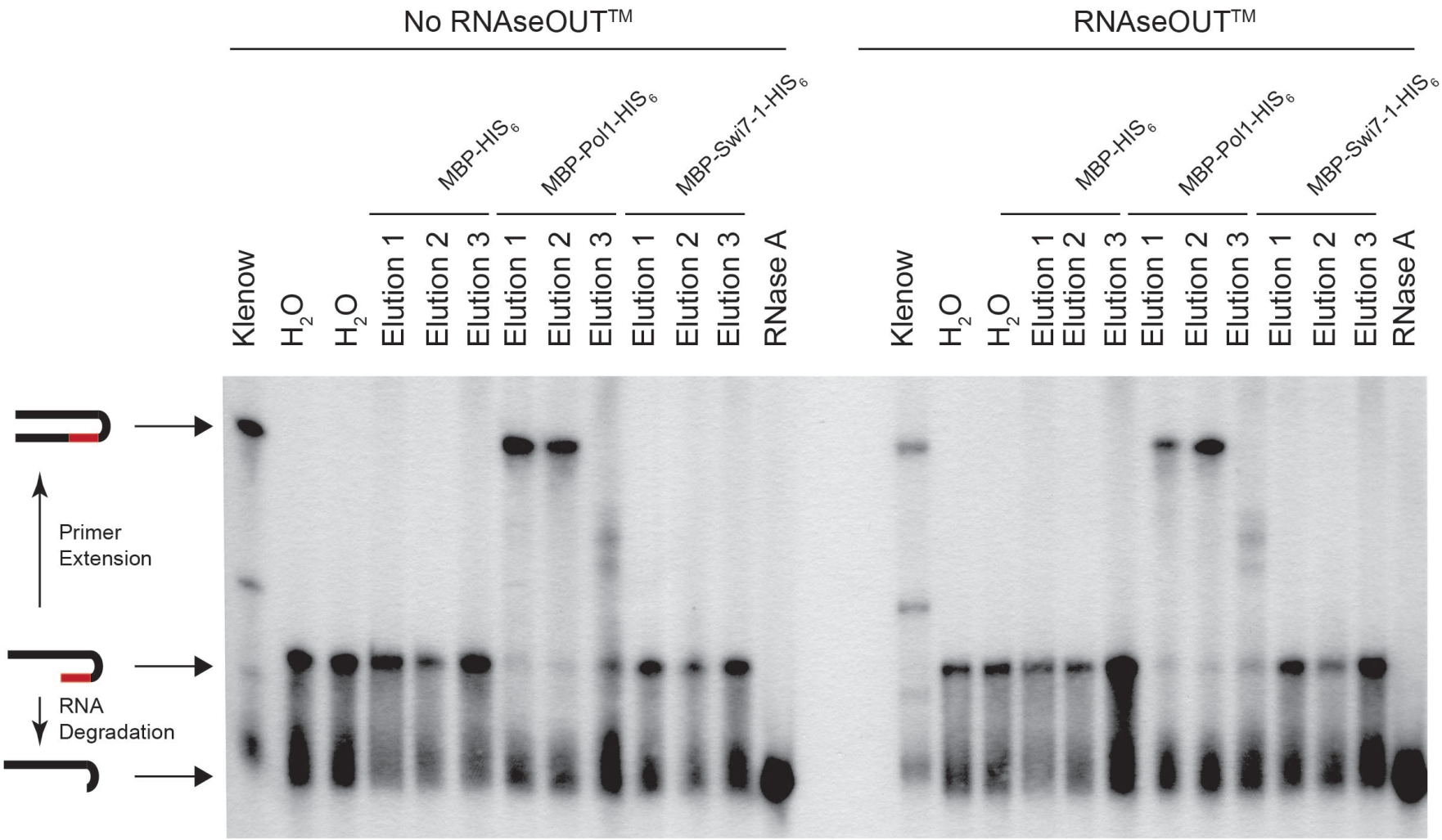

No RNAseOUT™
RNAseOUT™
Klenow
H2O
H2O
Elution 1
Elution 2
Elution 3
MBP-HIS6
MBP-Pol1-HIS6
MBP-Swi7-1-HIS6
RNase A
Primer Extension
RNA Degradation

Figure 3.5. Primer extension assay used to measure the polymerase activity of the purified recombinant proteins. The first three elution samples from the third purification column (Ni-NTA) for Pol1 (pRA5) and Swi7-1 (pRA6) were used to extend a hairpin substrate that has an RNA strand annealed to DNA, mimicking the natural substrate of Pol1. The reaction was carried out at 37°C for 1 h. The Klenow fragment (NEB) and water were used as positive and negative controls respectively. Purified samples from *E. coli* carrying an empty vector (that contains MBP-His_6) (pRA4) were also used as controls. Protein samples from *E. coli* carrying the pRA4 plasmid did not display any polymerase, nuclease or helicase activities, suggesting that our purification strategy was sufficiently stringent. The recombinant Pol1 was able to extend the hairpin, more competently than the Klenow fragment. However, the recombinant Swi7-1 protein was unable to extend the primer template within the duration of the reaction. RNAseOUT was used to inactivate ribonuclease activity whilst an excess of RNAse A was used to determine what effect contaminating ribonucleases would have on the reaction mixture.

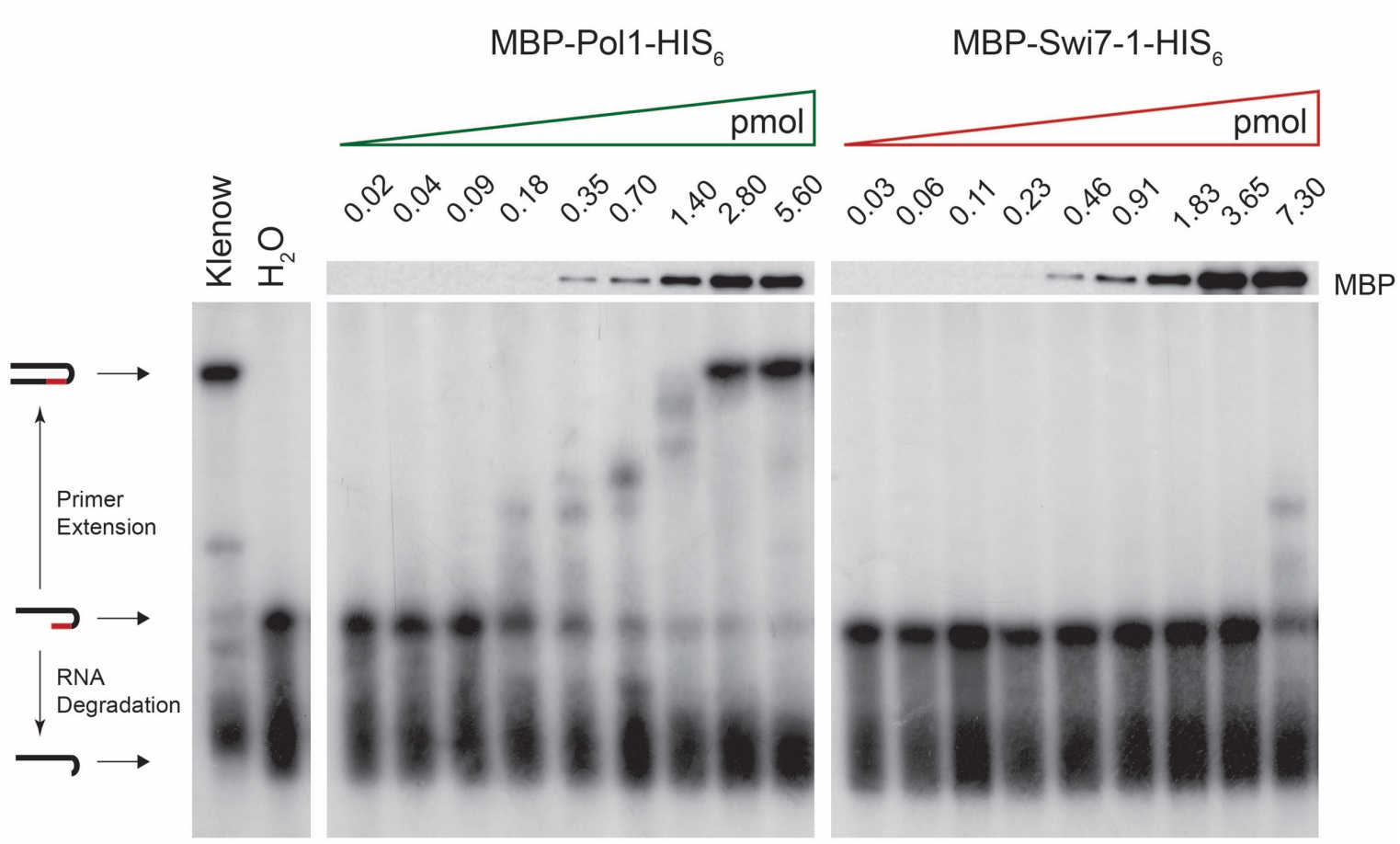

MBP-Pol1-HIS6
MBP-Swi7-1-HIS6
pmol
pmol
Klenow
H2O
0.02
0.04
0.09
0.18
0.35
0.70
1.40
2.80
5.60
0.03
0.06
0.11
0.23
0.46
0.91
1.83
3.65
7.30
MBP
Primer Extension
RNA Degradation

Figure 3.6. Under the experimental conditions used, the recombinant Swi7-1 protein has around 2.5-5% activity of its wildtype counterpart. This experiment is similar to the one presented in Figure 3.5 but with modifications. The reaction was still carried out at 37°C for 1 h but the protein samples purified from the strain carrying the empty vector (pRA4) were not used. Moreover, the levels of the recombinant Pol1 and Swi7-1 proteins were titrated up. The concentrations of the protein samples were determined using the BCA protein assay (Pierce) and the amounts of the proteins used for each reaction was calculated from this concentration. The relative quantities of the proteins used was also assayed via Western blotting. For Swi7-1, polymerase activity was detected only in the reaction mixture containing the highest amount of the protein (7.3 pmol). Given that 7.3 pmol of Swi7-1 could only extend the primer to the same extent as 0.18-0.35 pmol of Pol1 (under identical conditions), Swi7-1 is judged to have between 2.5-5% activity. The percentage activity was calculated by dividing the appropriate concentration of Pol1 (0.18 or 0.35 pmol) with the highest concentration of Swi7-1 used (7.3 pmol) (because they have comparable activities) multiplied by 100%.

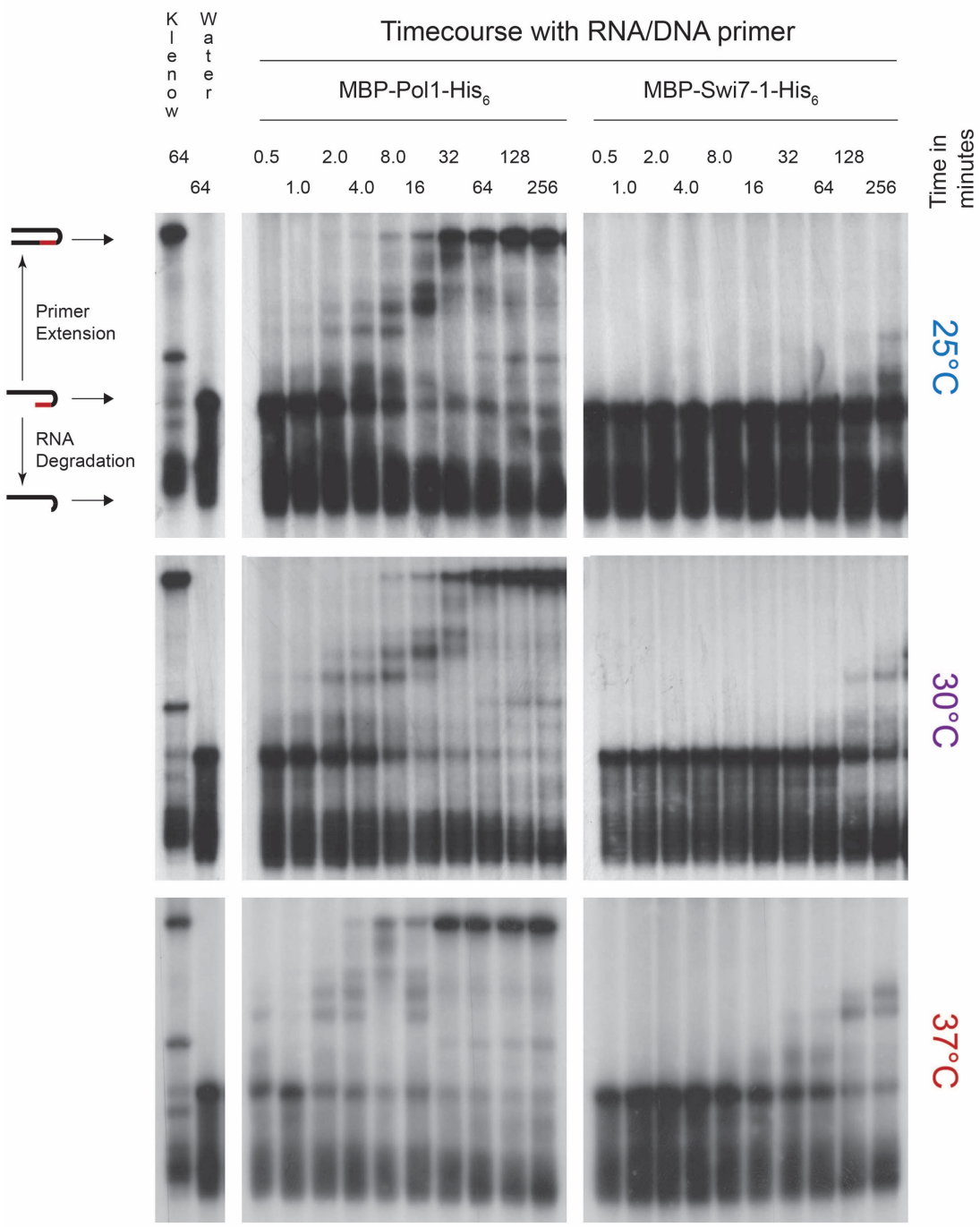


Figure 3.7. The primer extension assay was used to carry out time-course experiments at the physiological temperatures of fission yeast. 5.6 pmol of recombinant Pol1 and 7.3 pmol of Swi7-1 were used to extend the primer substrate at 24°C, 30°C and 37°C, temperatures that are routinely used to grow *Sz. pombe* strains. For each condition, parallel reactions were initiated at a single time and stopped at different time points.

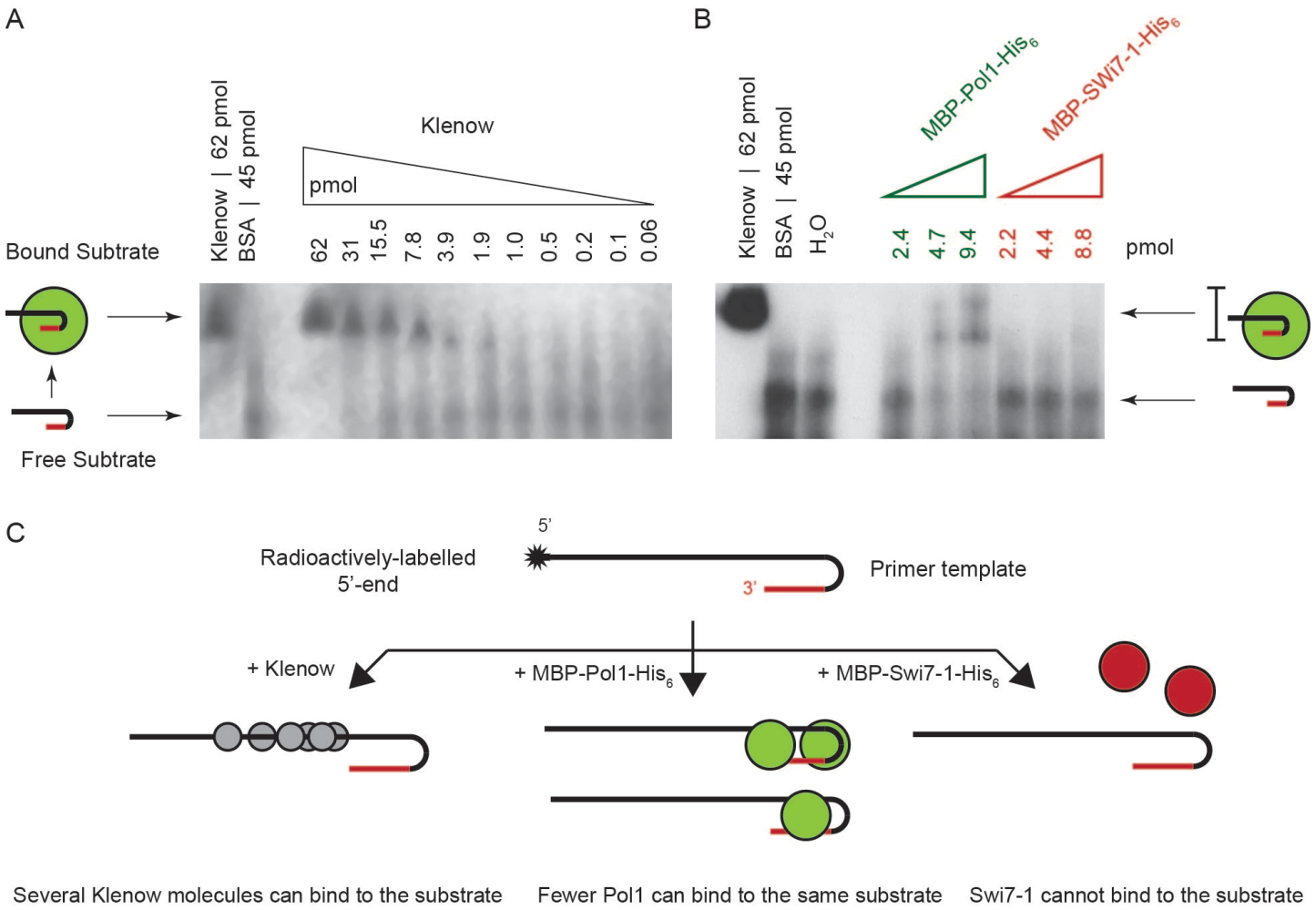

A
Klenow | 62 pmol
BSA | 45 pmol
Klenow
pmol
62
31
15.5
7.8
3.9
1.9
1.0
0.5
0.2
0.1
0.06
Bound Subtrate
Free Subtrate
B
Klenow | 62 pmol
BSA | 45 pmol
H2O
MBP-Pol1-His6
MBP-SWI7-1-His6
2.4
4.7
9.4
2.2
4.4
8.8
pmol
C
Radioactively-labelled
5'-end
5'
3'
Primer template
+ Klenow
+ MBP-Pol1-His6
+ MBP-Swi7-1-His6
Several Klenow molecules can bind to the substrate
Fewer Pol1 can bind to the same substrate
Swi7-1 cannot bind to the substrate

Figure 3.8. Gel shift assays reveal that Swi7-1 has a much reduced affinity for its template. This provides a rationale for the difference in polymerase activity in our primer extension assays. (A) The Klenow fragment was used as a positive control to determine whether the primer substrate used previously would be adequate as a substrate for gel shifts assays. At quantities as low as 1.9 pmol, the Klenow fragment was able to produce a shift, indicating nucleic acid-protein interaction. 45 pmol of BSA was not able to produce any shift. (B) At 4.7 pmol of Pol1, gel shifts were detectable whilst for Swi7-1, no shifts were detected even when using 8.8 pmol of the protein. (C) Summary of the difference between binding of the Klenow fragment and the recombinant Pol1 and Swi7-1 proteins to the substrate.

3.4 GENERATING NEW *SWI7* MUTANTS TO TEST THE RELATIONSHIP BETWEEN THE DNA BINDING AFFINITY OF POL1 AND ITS SWITCHING DEFECTS.

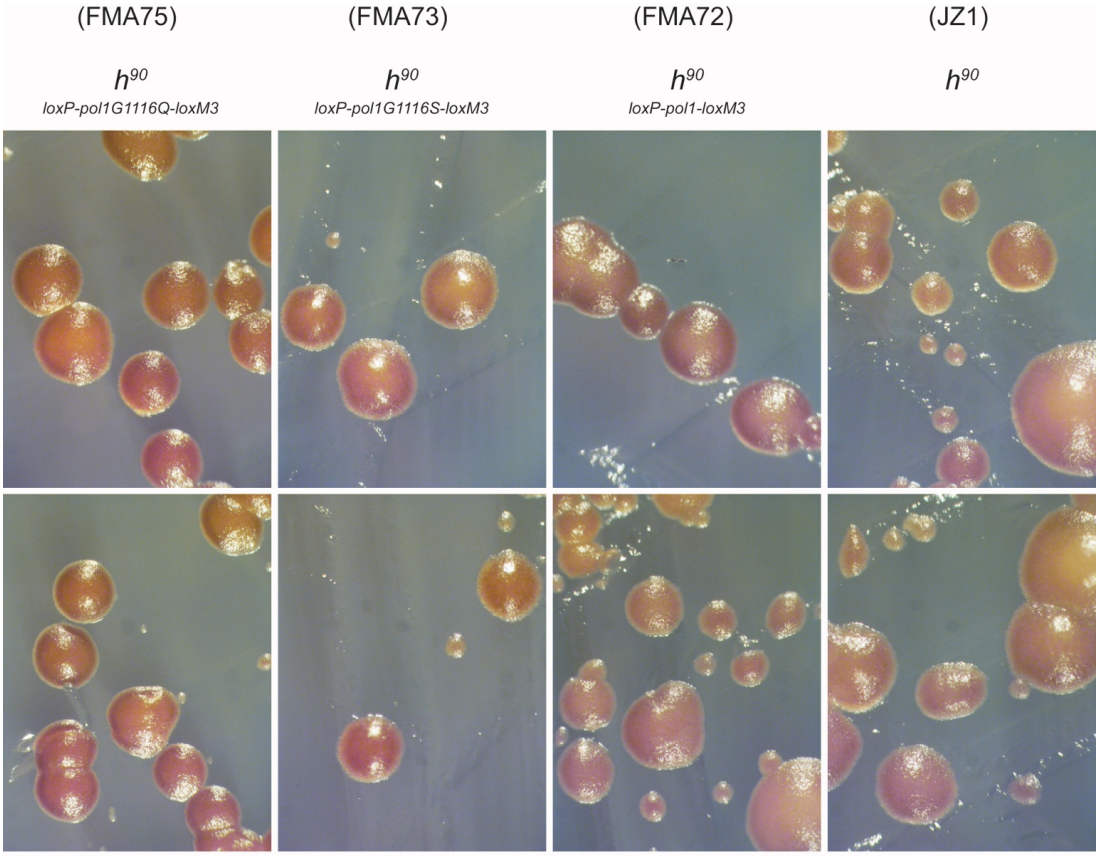
Results thus far suggest that the Swi7-1 protein has a reduced affinity for its substrates. We wanted to confirm whether this was also the case *in vivo*. We decided to adopt an indirect approach where we targeted the incriminating residue in the switching- and imprinting-defective *swi7-1* for mutagenesis. In particular, we wanted to generate mutants of *pol1* that were switching-defective, but to a lesser extent that *swi7-1*. For that purpose, we chose to mutate the Gly1116 residue in the wildtype allele to an aspartic residue. Aspartate and glutamate are comparable residues and both lose protons at neutral pH. However, aspartic acid also has a lesser potential for steric hindrance compared to glutamic acid. As such, Pol1G1116D would be predicted to have an affinity for its substrate that is intermediate between those of Pol1 and Swi7-1 (Pol1G1116E). As controls, we also aimed to mutate the Gly1116 residue to either serine (Ser1134 in Pol1 from *S. cerevisiae* aligns with Gly1116 in Pol1 from *Sz. pombe*) or glutamine (a residue that offers similar steric hindrance compared to glutamic acid, but is polar and not negatively charged). Both Ser1116 and Gln1116 could serve to strengthen the interaction between Pol1 and its substrate (Perera et al., 2013).

When generating mutant alleles, it is customary to delete the endogenous gene and use a vector expressing the mutant alleles to complement it. However, given that small differences in the levels of Pol1 could affect the read-out (iodine staining) of the mutants, an alternative approach was adopted. We chose to use a cassette-exchange method (Watson et al., 2008), coupled to a commercial site-directed mutagenesis kit (QuikChange, Agilent) to generate mutant alleles of *pol1* at the genomic loci.

The novel alleles carry *loxP* and *loxM3* scars. To ensure that these scars do not affect gene expression, *loxP-pol1-loxM3* and *loxP-swi7-1-loxM3* alleles that differ from *pol1* and *swi7-1* respectively only with respect to those scars

were also created. The novel alleles were tested for their efficiency to switch cell-types by iodine staining. Strains encoding the *loxP-pol1-loxM3* allele had an indistinguishable phenotype from strains carrying the wildtype *pol1* gene. Likewise, no phenotypic differences could be seen between cells carrying the *swi7-1* and the and *loxP-swi7-1-loxM3* alleles. This suggests that the *loxP* and *loxM3* scars have no physiological consequences. The novel *loxP-pol1G1116S-loxM3* and *loxP-pol1G1116Q-loxM3* alleles did not behave dissimilarly to wildtype upon staining, indicating that cells carrying these alleles form the Pol1-dependent imprint at wildtype levels and, thus, switch cell-type efficiently. On the other hand, the *loxP-pol1G1116D-loxM3* allele (henceforth referred to as *swi7-2*) had an intermediate phenotype between those of the wildtype and the *swi7-1* allele upon iodine staining (Fig 3.9), fitting the hypothesis that Pol1's interaction with its nucleic acid substrate is critical for stable imprinting at *MPS1*.

It has been reported that the *swi7-1* allele confers increased sensitivity to both HU and MMS, perhaps highlighting a role for Pol1 in alkylation damage repair (Koulintchenko et al., 2012). The hypersensitivity of cells carrying the *swi7-1* allele to both HU and MMS could be attributed to decreased affinity of the mutant Pol1 for its DNA/RNA substrates. We wanted to see whether cells carrying the *swi7-2* allele were also sensitive to both HU and MMS. We found that *sen1-2* did confer increased sensitivity to both HU and MMS but, importantly, cells carrying the *swi7-2* allele were less sensitive to both HU and MMS than cells with the *swi7-1* allele (Fig 3.10). Should the Pol1-dependent sensitivity to HU and MMS for **this class of mutants** inversely correlate with the affinity of the enzyme for its substrates, these results fits with the hypothesis that Swi7-2 has an intermediate affinity for its DNA/RNA substrate between those of Pol1 and Swi7-1. However, the binding affinity of Swi7-2 compared to that of Pol1 and Swi7-1 needs to be tested empirically (e.g by using EMSA assays). It also supports the notion that Pol1's affinity for its substrates is important for the imprinting process at *MPS1*.


(FMA75)
h^{90}
loxP-pol1G1116Q-loxM3
(FMA73)
h^{90}
loxP-pol1G1116S-loxM3
(FMA72)
h^{90}
loxP-pol1-loxM3
(JZ1)
h^{90}

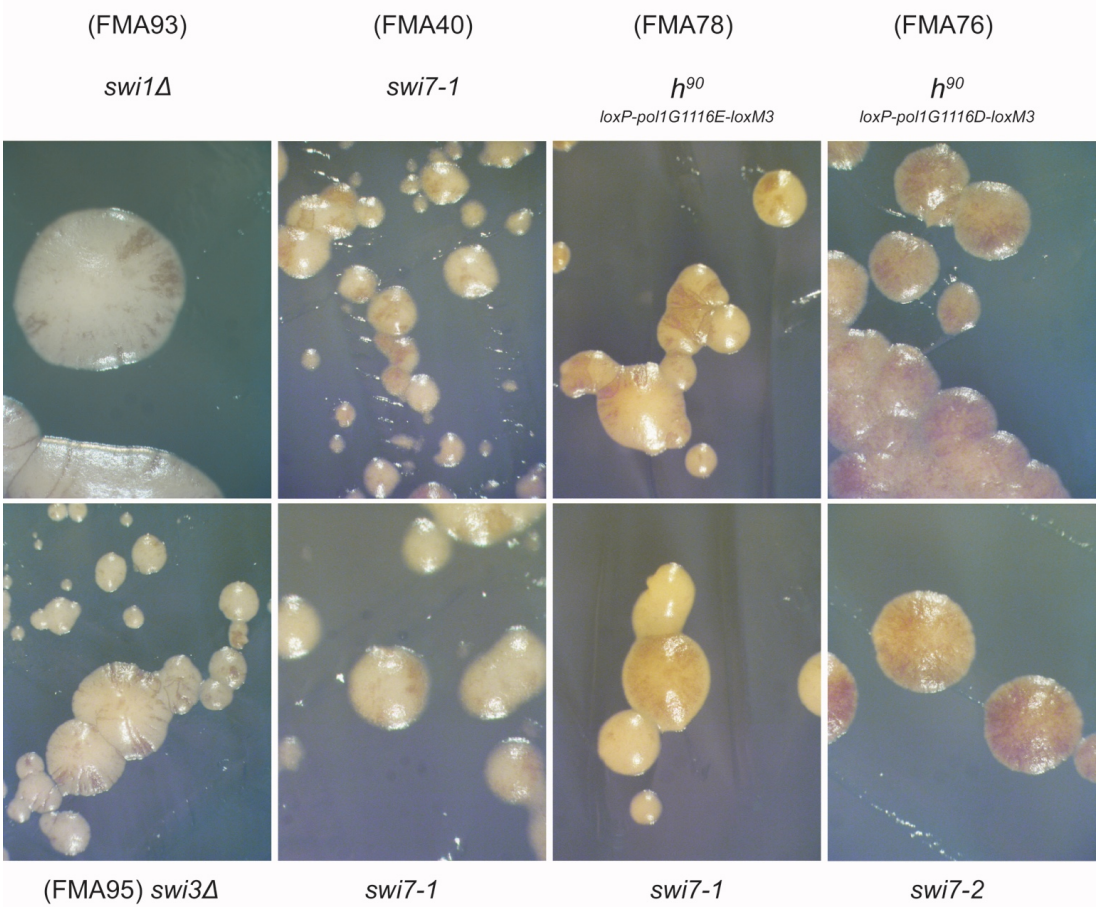

(FMA93)
swi1Δ
(FMA40)
swi7-1
(FMA78)
h^{90}
loxP-pol1G1116E-loxM3
(FMA76)
h^{90}
loxP-pol1G1116D-loxM3
(FMA95) swi3Δ
swi7-1
swi7-1
swi7-2

Figure 3.9. Iodine staining demonstrates that the novel *swi7-2* (*pol1G1116D*) mutant allele is switching-defective but to a lower extant than *swi7-1* (*pol1G1116E*). Iodine was used to first assess whether the *loxP* and *loxM3* scars occurring on either side of the *pol1* gene impact switching efficiency. This was found not to be the case. Iodine staining was then used to determine whether the novel alleles of *pol1* alleles were defective for switching. The *loxP-pol1G1116S-loxM3* and *loxP-pol1G1116Q-loxM3* alleles behaved similarly to wildtype upon staining. The third mutant allele, *pol1G1116D* was found to be switching-defective, but to a lower extent that *swi7-1*. Consequently, this allele was then renamed *swi7-2*.

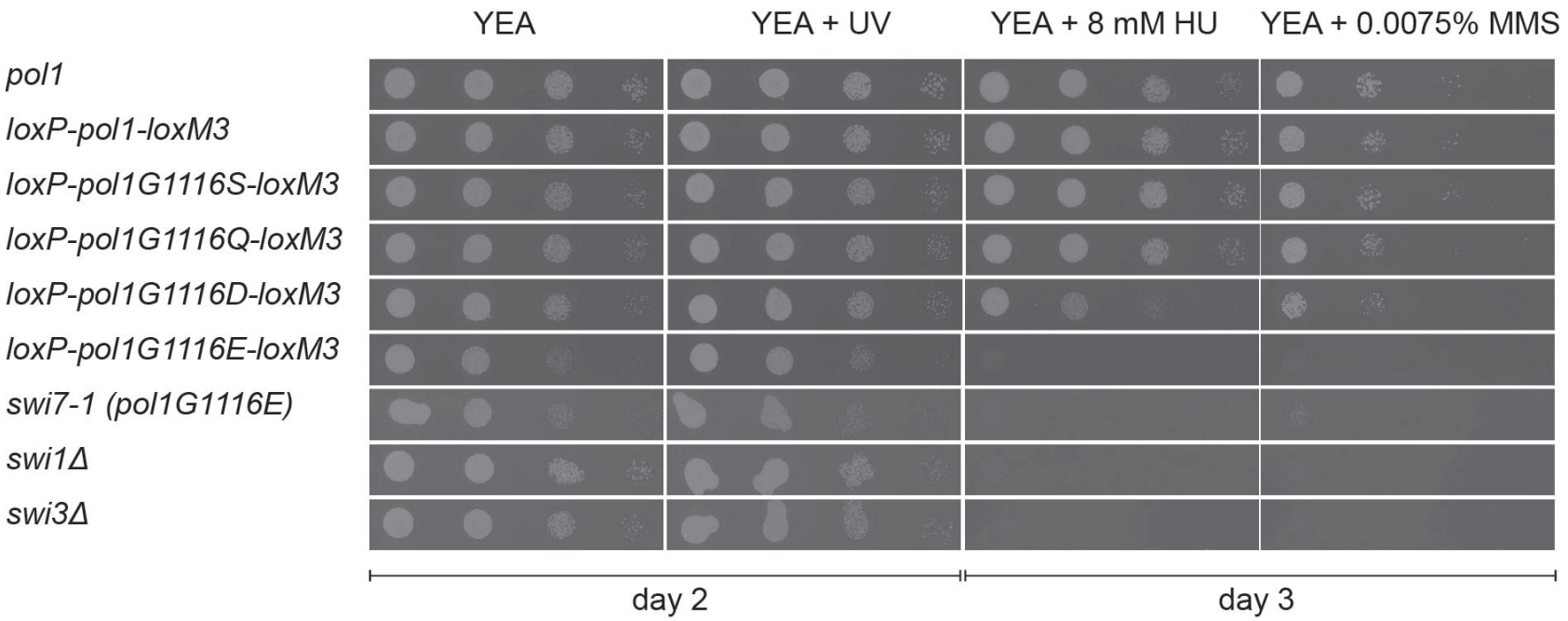


Figure 3.10. The *swi7-2* allele has an intermediate phenotype between those of *pol1* and *swi7-1* in relation to sensitivity to HU and MMS. Fission yeast strains were grown from frozen for 2 days at 33°C. The cells were then diluted in water to obtain cell suspensions of 5 x10^6, 5 x 10^5, 5 x 10^4 and 5 x 10^3 cells/ml. 10 µl of each suspension was pipetted onto either non-selective medium (with or without exposure to UV) or non-selective medium supplemented with either HU or MMS. Cells were grown up to 4 days and imaged daily. Strains used (top to bottom): JZ**1**, FMA**72**, FMA**73**, FMA**75**, FMA**76**, FMA**78**, FMA**40**, FMA**93** and FMA**95**.

3.5 A MUTANT OF *SPP1*^{PRII}, APPARENTLY DEFECTIVE FOR CATALYSIS, AND DELETION MUTANTS OF *MCL1*^{CTF4} LEAD TO DEFECTS IN CELL-TYPE SWITCHING

Given the interaction between Pol1 and its substrate seem to be a requisite for stable imprinting at *MPS1*, we wondered whether the binding partner of Pol1, Mcl1^{Ctf4}, was also required for the stability of the imprint at *MPS1*. Mcl1 is known to tether Pol1 to the rest of the replisome. We first tested whether *mcl1* was required for cell-type switching. We acquired *mcl1*Δ mutants from two sources: (i) from the Bioneer deletion library and (ii) from Prof Takashi Toda (Mamnun et al., 2006) and crossed them with the *h*^{90LEU2} homothallic marker. Both *mcl1*Δ alleles were marked with the kanamycin reporter cassette and the homothallic marker was marked with the *LEU2* reporter gene from *S. cerevisiae*. After crossing, random spores were allowed to germinate and haploids with both the kanamycin and *LEU2* marker were selected for. Two independent clones from each cross that carried both the *mcl1*Δ and *h*^{90LEU2} alleles were grown on sporulation medium for two days at 30°C. After sporulation, the plates were stained with iodine vapour and imaged. Colonies carrying the *h*^{90LEU2} and *mcl1*Δ^{Toda} alleles (FMA146) stained darkly, without the speckled/mottled phenotype characteristic of switching-defective mutants. However, some colonies stained lightly (as seen with some *pol1-1* colonies), suggesting a stochastic loss of switching. Meanwhile, some colonies of cells carrying the *h*^{90LEU2} and *mcl1*Δ^{Bioneer} alleles (FMA155) also stained lightly. However, the remaining colonies stained less darkly than wildtype. As such, this strain is characterized by a majority of cells switching cell-type inefficiently with some cells unable to switch at all.

The difference between the *mcl1*Δ^{Toda} and *mcl1*Δ^{Bioneer} alleles could reflect the different strategies used to generate the different mutants. We tested cells carrying either of the *mcl1*Δ alleles by PCR and confirmed that the *mcl1* gene was either completely deleted or deleted by disruption. We note however that *mcl1*Δ has been shown to be both viable (Mamnun et al., 2006) or recessive lethal (Williams and McIntosh, 2002). These data can be reconciled by the

presence of unaccounted for synthetic defects or suppression. Perhaps, the difference between the phenotypes observed for *mcl1*Δ^{Toda} and *mcl1*Δ^{Bioneer} can also be accounted for by some unknown genetic interaction. Common to these two alleles, however, are colonies that do not stain with iodine vapour at all. This phenotype is dissimilar to those of cells that are defective in imprinting such as *swi7-1*.

The imprint at *MPS1* has been shown to be made up of two ribonucleotides (Vengrova and Dalgaard, 2004, Vengrova and Dalgaard, 2006) (Fig 3.12). It seems likely then that the primase subunit of Polymerase α (Spp1^{Pri1}) has a role in imprinting at the *MPS1* locus. Importantly, *spp1*Δ is lethal and no cell-type switching- or imprinting-defective mutants of *spp1* have yet been identified. In the present study, we have noticed that a tagged allele of *swi7* (*swi7-1-TAP*) but not *pol1-TAP* confers cold-sensitivity (not shown). It is known that a conserved motif at the extreme C-terminus of Pol1 is required for interaction with the primase subunits of Polymerase α (Kilkenny et al., 2012). The *swi7-1-TAP* allele then may be defective for interaction with primase. Interestingly, we also noticed that a tagged allele of *spp1* (*spp1-GFP*) (Yang et al., 2005) confers a similar cold-sensitivity presumably through having impaired enzymatic ability. We wondered whether these two alleles also shared switching-defects. To do so, the *spp1-GFP* was crossed in cells carrying the *h*^{90LEU2} marker and we assessed the ability of the resulting cells to switch cell-type and hence to stain with iodine vapour. Colonies carrying the *spp1-GFP* allele stained lightly, similar to heterothallic strains (Fig 3.11). This could indicate that the primase subunit of Polymerase α is required for the imprinting at *MPS1* and that the GFP tag interferes with this process. To ensure that the phenotype seen was specific to the *spp1-GFP* allele and not some general result of tagging replisome components with some bulky fluorescent protein, we also tested the ability of cells carrying both the *psf2-YFP* (Yang et al., 2005) and *h*^{90LEU2} alleles to switch cell-types. Unlike *spp1-GFP* colonies, *psf2-YFP* colonies stained darkly suggesting that the *spp1-GFP* allele has specific defects in cell-type switching. It is interesting to note that *spp1-YFP* behaves like a heterothallic strain and not like switching-defective

mutants such as *swi1*Δ or *swi7-1*. This suggests a stronger imprinting-defect in cells carrying the *spp1-GFP* allele.

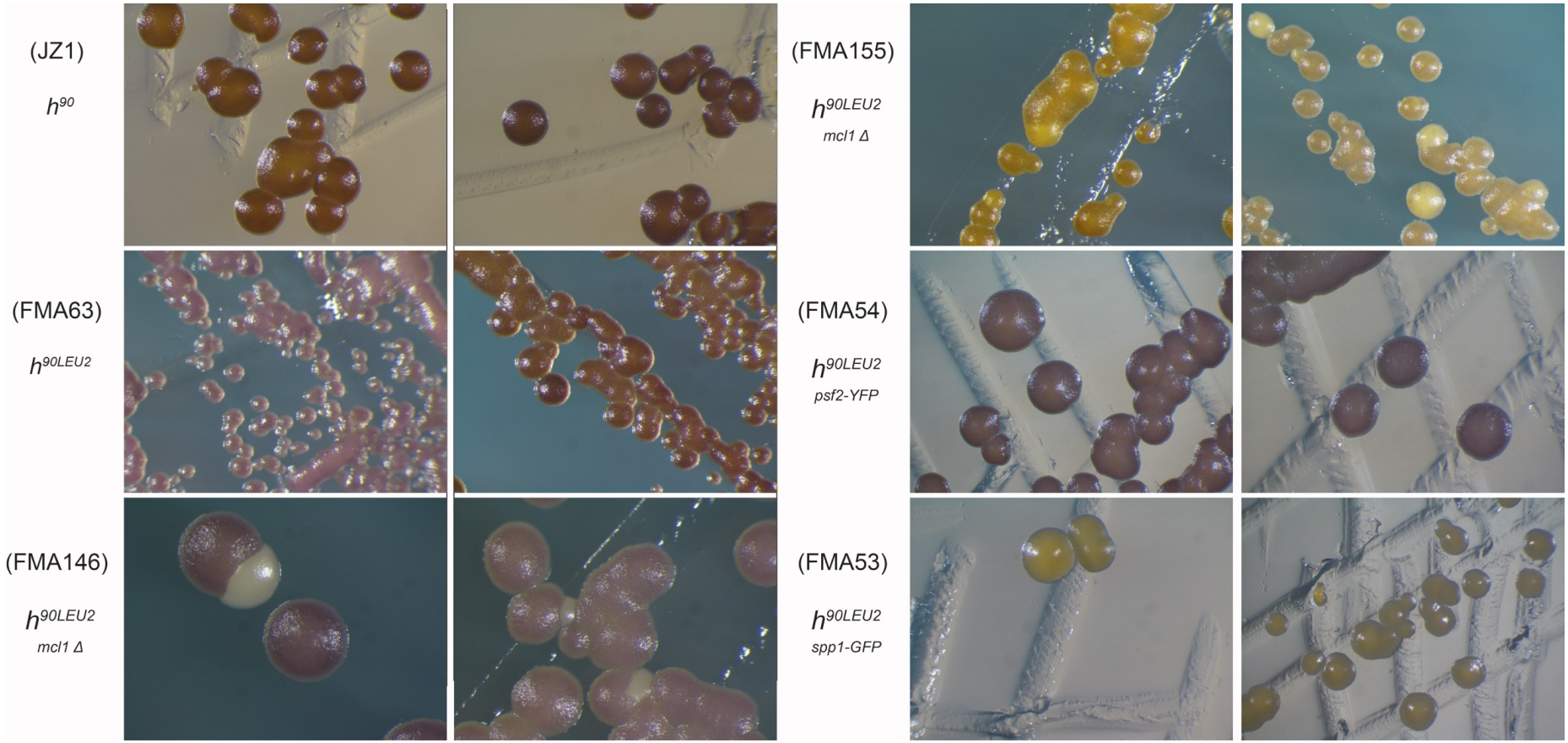


Figure 3.11. Iodine staining demonstrates that both the *mcl1*Δ and *spp1-GFP* alleles present some defects in cell-type switching. Iodine staining was used to assess defects in cell-type switching in strains carrying either the *mcl1Δ* or *spp1-GFP* allele. A wildtype strain and a strain carrying the *psf2-YFP* allele were used as controls.

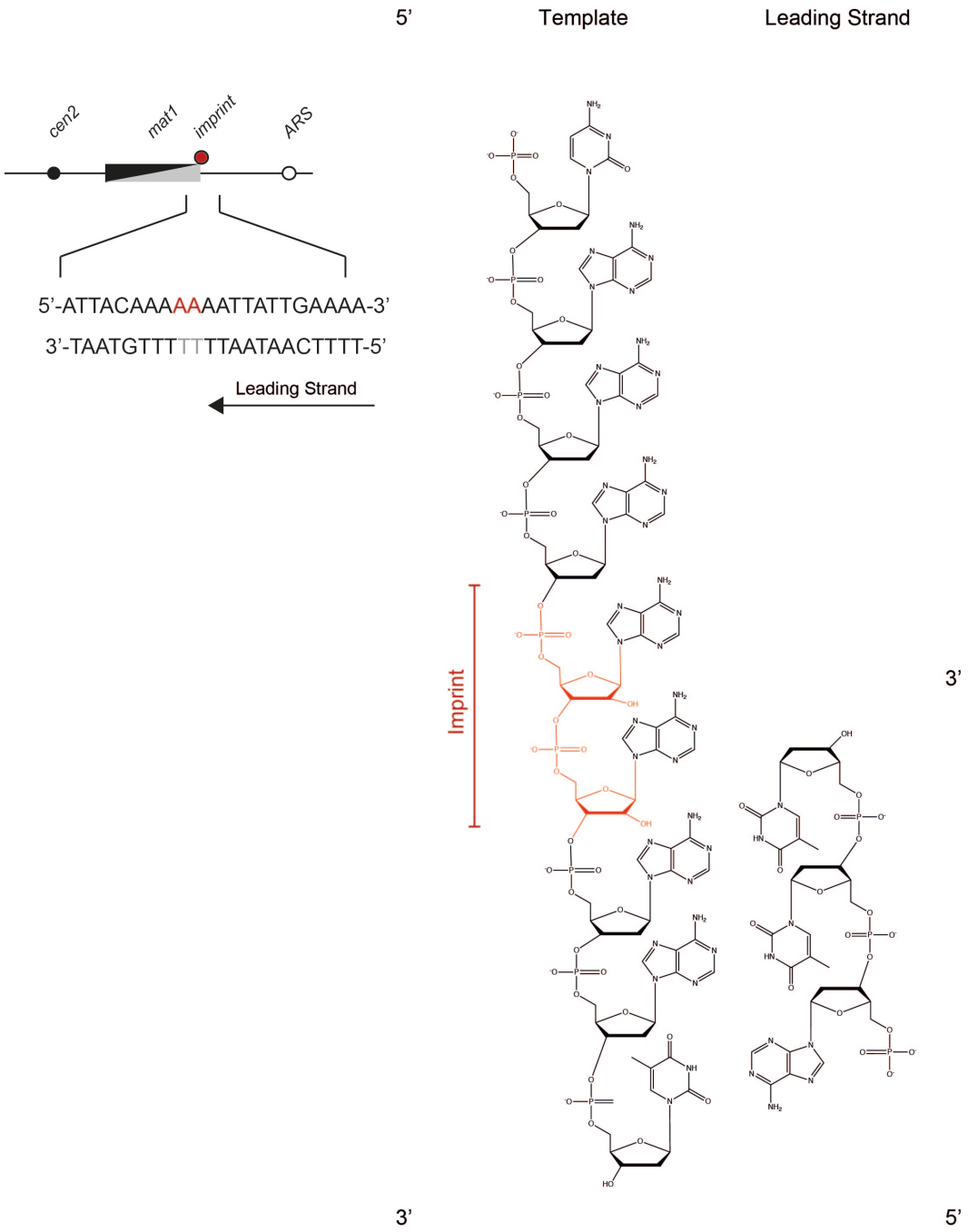


Figure 3. 12. The imprint is made up of two ribonucleotides. It is likely that the switching-deficient phenotypes associated with *mcl1Δ, swi7-2* and *spp1-GFP* are linked to inefficient imprinting.

CHAPTER 4: UNRAVELLING THE ROLE OF SEN1 AS A MODULATOR AT THE INTERFACE OF DNA REPLICATION AND TRANSCRIPTION

4.1 BACKGROUND

The DNA double helix is the common substrate of both DNA replication and transcription. Temporal overlap of the two processes is possible as, unlike DNA replication that is confined to the S phase, transcription occurs throughout the cell cycle. This can lead to collisions between the transcription and replication machineries (Aguilera and García-Muse, 2012). Head-on collisions between those two processes are more problematic than rear-end (or co-directional) collisions (Hamperl et al., 2017, Prado and Aguilera, 2005). In addition, transcription produces an intermediate that antagonizes the replication process. This intermediate, known as the R-loop, is a three-stranded nucleic species composed of the nascent mRNA bound to the template strand with the non-template strand displaced. R-loops are thermodynamically more stable than the equivalent dsDNA by virtue of the hybrid interaction between RNA and DNA molecules (Lesnik and Freier, 1995, Roberts and Crothers, 1992) and need to be removed through the activity of specific enzymes such as the RNase H endonucleases.

Unresolved collisions between the DNA replication and transcription machineries have been shown to trigger genetic instability in a wide array of organism, from *S. cerevisiae* (Wellinger et al., 2006) to B-lymphocytes (Barlow et al., 2013). Given the short window of time during which chromosomes need to be fully and faithfully duplicated and the dire consequences of failing to complete DNA replication in a timely fashion, it has been hypothesized that cells have evolved mechanisms to avert or quickly resolve such conflicts by forcibly disengaging the transcription machinery from DNA, favouring fork progression. This would cause premature transcription termination (transcription attenuation). Such a model dictates that enzymes capable of

resolving R-loops can be quickly enriched at sites of collisions between replication forks and transcribing RNA polymerases (or R-loops). Alternatively, these enzymes might travel with either the replication or transcription machinery allowing efficient recruitment at sites of collisions. In the model organism *Saccharomyces cerevisiae*, a possible candidate for this role is the DNA/RNA helicase, Sen1. The latter participates in premature transcription termination although the precise mechanism through which it does so has not been fully characterized yet. It is known that Nab3 and Nrd1, that form a hetero-trimer with Sen1, bind to nascent mRNA molecules (Jamonnak et al., 2011) whilst full-length Sen1 is able to disengage the transcription machinery in a manner reminiscent of the bacterial Rho termination factor *in vitro* (Porrua and Libri, 2013). Sen1 also associates with forks (Alzu et al., 2012). Sen1, then, is an ideal candidate to rid the genome of R-loops at actively transcribing loci in favour of replication forks. Such a mechanism would help to resolve problematic collisions between forks and transcribing complexes.

This project originated from the observation that Sen1 co-precipitates with the replisome in a mass spectrometric screen of S phase GINS IP (De Piccoli, unpublished data). Here, I have sought to characterize this interaction and to investigate its biological relevance. I have found that Sen1 interacts with the replisome by virtue of its N-terminal domain and one interactor of Sen1 (Ctf4) has been identified. By screening point mutants of *SEN1*, one mutant that is unable to bind to the replisome has been isolated. This mutant is healthy in isolation but is synthetically lethal in the absence of both RNase H1 and H2.

4.2 SEN1 CO-PRECIPITATES WITH REPLISOME COMPONENTS IN S PHASE

Mass spectrometric screens are highly sensitive and thus prone to misidentification of peptides as false-positives. In order to validate Sen1's ability to co-precipitate with GINS, Sen1 was C-terminally tagged with 9MYC and crossed with a strain carrying the *TAP-SLD5* allele (Sld5 being a component of the GINS complex). We grew the cells at 24°C and harvested them either arrested in G_1 or synchronized in S phase. These samples were then used for GINS IPs followed by Western blotting. Pierce Universal nuclease was added to a final concentration of 1.6 U/µl to remove any DNA or RNA molecules within the extracts. In G_1, Sld5 is known to interact with Ctf4 but not to other non-GINS components of the replisome. At the onset of S phase, both Cdc45 and GINS are recruited to origins where the MCM2-7 double hexamer was already loaded in G_1. This leads to the formation of the CMG helicase that is required for complete formation of the replisome (Fig 4.1). We found that, in G_1, Ctf4 but neither Cdc45 nor any MCMs co-precipitated with GINS (Fig 4.2), fitting the established pattern of CMG assembly (Gambus et al., 2009).

Sen1 also did not co-precipitate with the GINS in G_1. In S phase, however, several components of the replisome co-precipitated with GINS. Moreover, Sen1 also co-precipitated with GINS, validating the results from the earlier mass spectrometric screen (De Piccoli, unpublished data) (Fig 4.2). This suggests that Sen1 interacts with the replisome after assembly in S phase but does not interact with GINS directly in G_1. We then used a strain encoding a *SEN1-TAP* allele to immuno-precipitate Sen1 in G_1, S and G_2. Replisome components co-precipitated with Sen1 only in S phase, suggesting that Sen1 forms an integral part of the replisome (Fig 4.3). Tellingly, none of the proteins we probed against co-precipitated with Sen1 outside of S phase. Whilst this could be indicative of an absence of physical interaction between Sen1 and the proteins probed for, an alternative interpretation is that the interaction between Sen1 and its binding partner(s) is strictly cell cycle dependent.

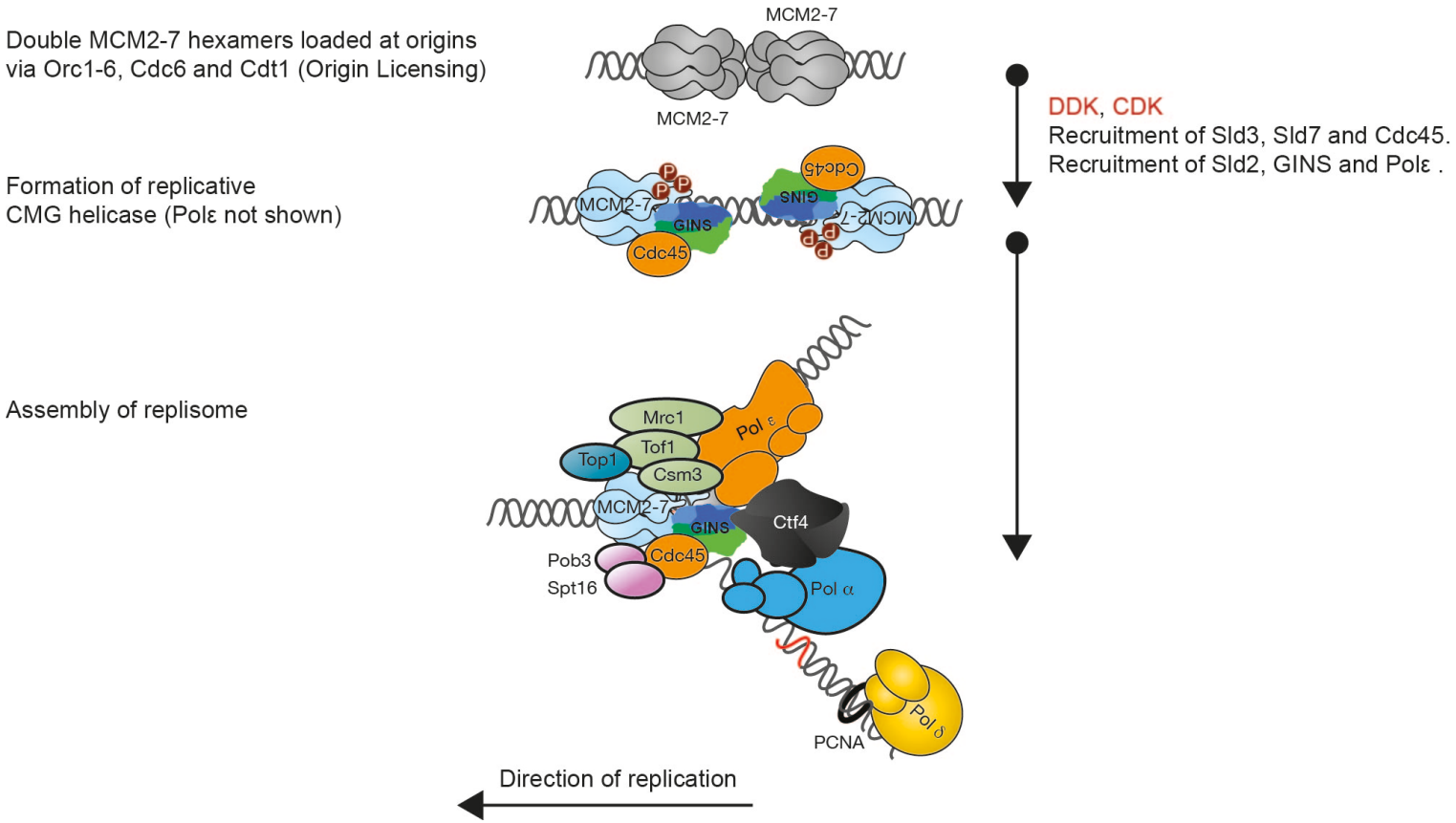


Figure 4.1. Sequence of events leading to CMG assembly at the onset of S phase. During origin-licensing two MCM2-7 hexamers are recruited sequentially at origins in G_1 by the concerted efforts of ORC1-6, Cdc6 and Cdt1. At this point in the cell cycle, GINS (Sld5, Psf1, Psf2 and Psf3) cannot interact with MCM2-7. It is only at the onset of S phase that Cdc45 and GINS are recruited to origins, forming the replicative CMG helicase at the centre of the eukaryotic replisome.

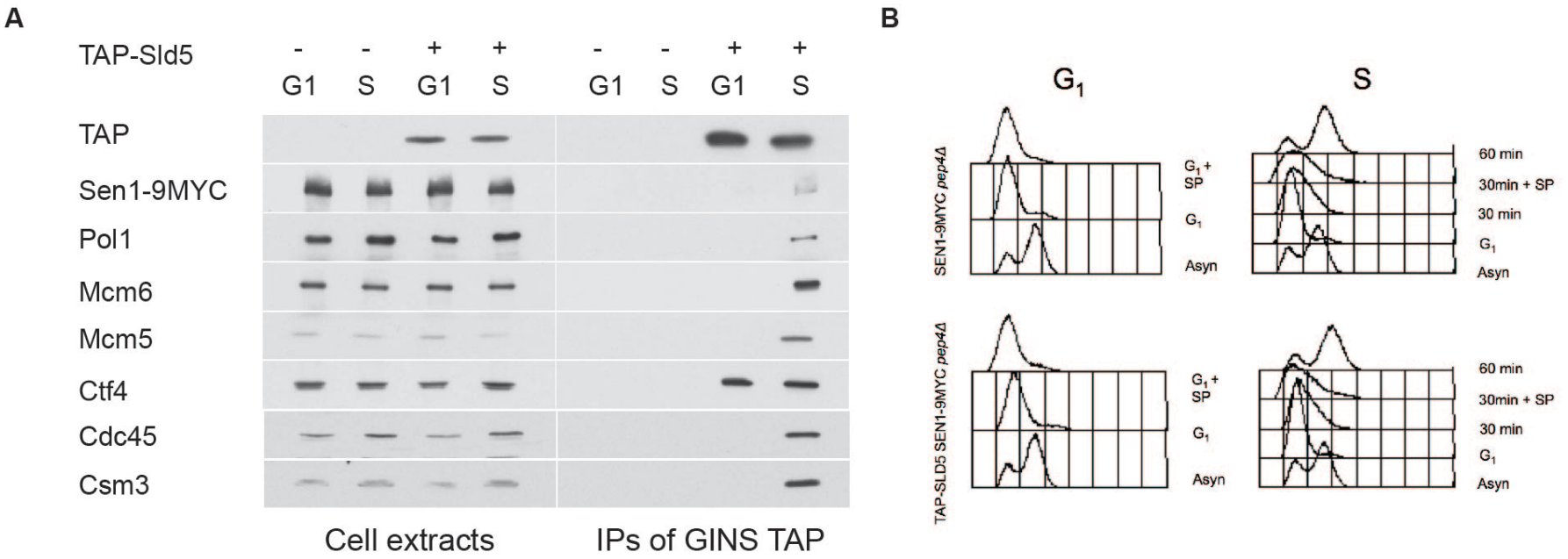


Figure 4.2. Sen1 interacts with components of the replisome in S phase but not in G_1. Cells carrying the *SEN1-9MYC* allele with or without the *TAP-SLD5* allele were grown in YPD at 24°C to a density of 0.7 x 10^7 cells/ml in 1 l cultures. The cells were arrested in G_1 using α-factor to a final concentration of 7.5 μg/ml for 3 h. After successful arrest (> 90% of cells shmooing and/or unbudded), the cells were either harvested and sample-prepared as described in Section 2.4.4 or the cells were released from α-factor arrest by washing twice in fresh YPD, re-suspended in 1 l YPD and released in S phase at 24°C. The cells were harvested 30 min post-release and sample-prepared. FACS samples were taken for asynchronous cultures (Asyn), at G_1 arrest (G_1), after preparation of the G_1 sample (G_1 + SP), at point of harvest in S phase (30 min), after preparation of the S phase sample (30 min +SP) and 30 min after S phase harvest (60 min). The samples were used for IPs using TAP beads. **(A)** Westerns of TAP, MYC and several replisome components for both cell extracts and TAP IPs. **(B)** FACS samples of the experiment. Strains used in this experiment: CS**1125** and CS**1126**.

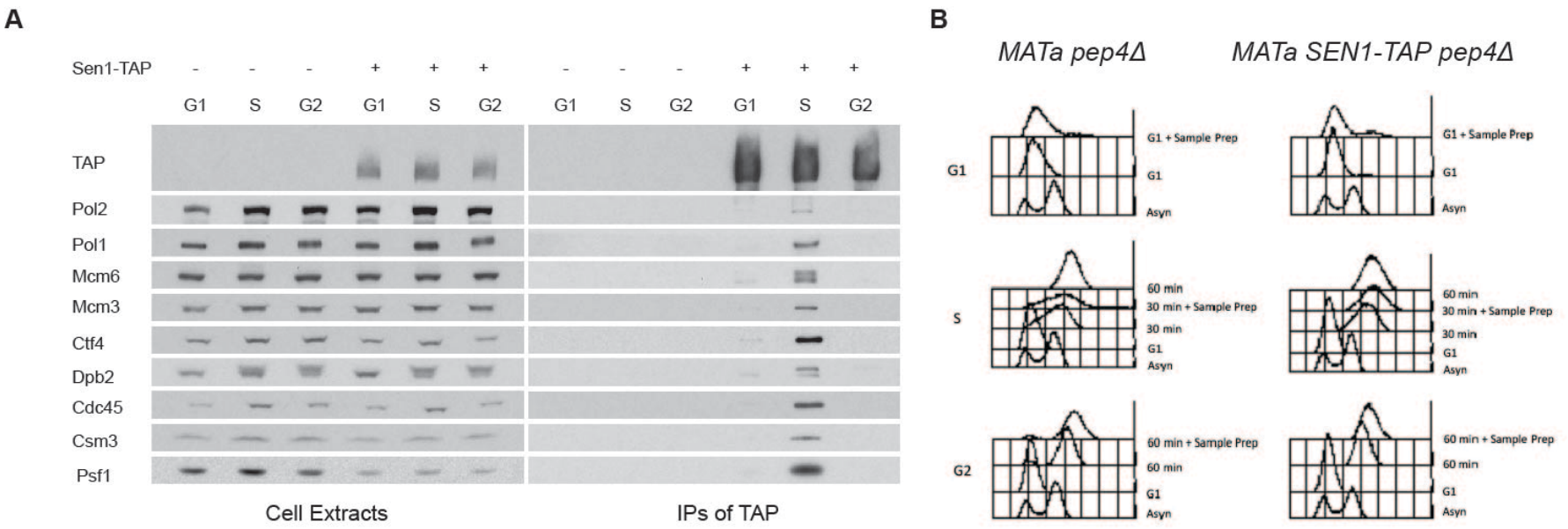


Figure 4 3. Sen1 interacts with several sub-complexes of the replisome in S phase. Cells carrying the *SEN1-TAP* allele were grown in YPD at 24°C to a density of 0.7 x 10^7 cells/ml in 1 l cultures. The cells were arrested in G_1 using α-factor to a final concentration of 7.5 μg/ml for 3 h. After successful arrest, the cells were either harvested and sample-prepared or the cells were released from α-factor arrest by washing twice in fresh YPD, re-suspended in 1 l YPD and released in S phase at 24°C. The cells were then harvested and sample-prepared either 30 or 60 min after release in S phase. FACS samples were taken for asynchronous cultures (Asyn), at G_1 arrest (G_1), after preparation of the G_1 (G_1 + Sample Prep) sample, at point of harvest in S phase (30 min), after preparation of the S phase sample (30 min +Sample Prep), at point of harvest in G_2 (60 min) and after preparation of the G_2 phase sample (60 min +Sample Prep). The samples were used for IPs using TAP beads. **(A).** Westerns of TAP and several replisome components for both cell extracts and TAP IPs. **(B).** FACS samples of the experiment. Strains used in this experiment: CS**74** and CS**1353**.

4.3 THE N-TERMINAL DOMAIN OF SEN1 IS REQUIRED FOR INTERACTION WITH THE REPLISOME

Having established that Sen1 interacts with replisome components, we then sought to determine what domain of the protein is required for this interaction. Sen1 is made of two domains, both of which are conserved albeit to different extents (Fig 4.4). Residues 1327-1656 encode a helicase domain that has been shown to be essential for viability (DeMarini et al., 1992). The *sen1-1* allele (Fig S.2) encodes a substitution in this helicase domain that reduces the extent of ATP hydrolysis and transcription termination *in vitro* (Porrua and Libri, 2013). Residues 106-859 encode an N-terminal domain of which Sen1 is the prototype. It is required for several protein-protein reactions, including with Rad2 (Ursic et al., 2004). Deletion of this domain is not lethal but its ablation leads to dysregulation of gene expression, shorter chronological life span and loss of mitochondrial DNA (Sariki et al., 2016).

We cloned N-terminally TAP-tagged constructs of Sen1 (Fig S.2) that spanned either one or both of these domains under the strong inducible *GAL1* promoter. We also took into consideration the predicted location of unstructured stretches of amino acids to determine the ends of the fragments (Fig 4.5). These fragments were cloned at an ectopic site in the genome at the *LEU2* locus on chromosome III. Strains carrying the different fragments were grown in YP-Raff to prevent premature expression of the constructs. When the required cell density was reached (0.7×10^7 cells/ml), the cells were synchronized in G_1 using α-factor and the medium was then switched to YPGAL, still in the presence of α-factor, triggering gene expression of the constructs for 35 min. The α-factor was removed by repeated washing with fresh YPGAL and cells were released in S phase in YPGAL at 24°C. Samples were then harvested in S phase (30 min after release at 24°C).

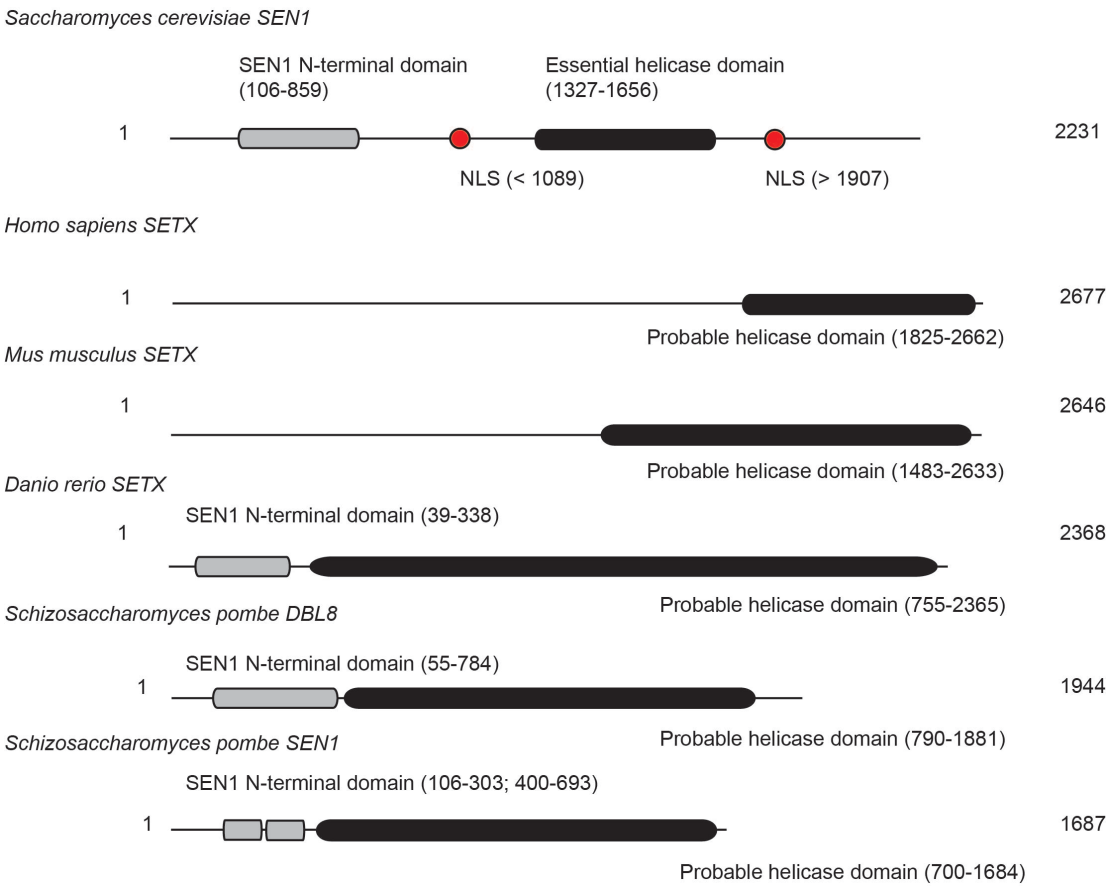


Figure 4.4. Sen1 is made up of two domains that are conserved in eukaryotes. Sen1 is made up of the Sen1 N-terminal domain at its N-terminus and a helicase domain at its C-terminus. These two domains are conserved in Sen1 orthologues although the Sen1 N-terminal domain is conserved to a lower extent, perhaps because it is not necessary for viability (Adapted from Interpro).

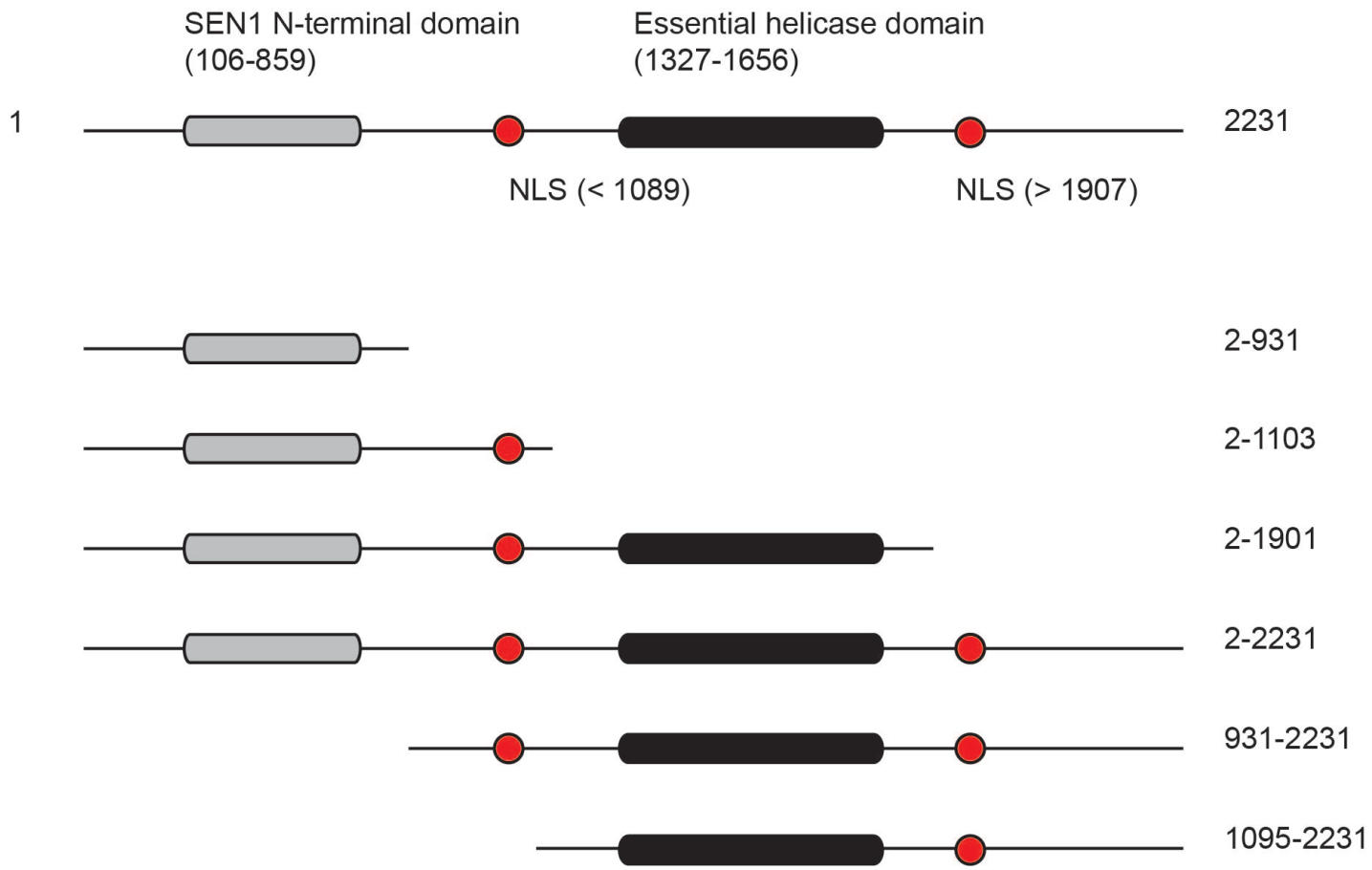


Figure 4.5. Schematic of Sen1 Constructs used to determine the domain through which the protein interacts with the replisome. Constructs of N-terminally TAP-tagged Sen1 fragments were cloned ectopically at the *LEU2* locus under control of the strong, inducible *GAL1* promoter.

This experiment revealed that fragments expressing the essential helicase domain were prone to quick degradation (Section 5.4). Why this is the case is uncertain but is congruent with the difficulty to overexpress the full-length protein in yeast (DeMarini et al., 1995). This also suggests the presence of some regulatory mechanism that ensures levels of Sen1 do not exceed a certain threshold. In order to achieve appreciable levels of the constructs expressing the essential helicase domain, we attempted to scale up the amount of yeast harvested from 1.75 x 10^9 cells to 7.0 x 10^9 cells. In addition, the cells were grown in YPGAL throughout the experiment to allow for continuous expression of the constructs. With these modifications, we were able to detect appreciable levels of Sen1 constructs expressing the helicase domain.

Immunoprecipitating the different constructs from cells harvested in S phase reveals that replisome components co-purified with constructs expressing the Sen1 N-terminal domain (Fig 4.6). To control for the possibility that the N-terminal domain of Sen1 might interact non-specifically with the TAP beads used for immuno-precipitation, two isogenic yeast strains were generated with the *GAL1-3HA-SEN1 (2-931)* allele and with or without *TAP-MCM3*. The cells were harvested in S phase after expression of the *GAL1-3HA-SEN1 (2-931)* allele and these samples were then used to generate cell extracts from which TAP-tagged proteins were immuno-precipitated. Results from this experiment confirmed that the N-terminal domain of Sen1 does not interact non-specifically with TAP beads. To ensure that the constructs expressing the helicase domain were folding correctly, we crossed the different fragments in a strain carrying the *td-MYC-sen1-1* and *GAL1-UBR1* genes and tested the ability of the different constructs to complement the temperature sensitivity of the *td-MYC-sen1-1* allele at the non-permissive temperatures (37°C) in the presence of galactose. The *td-MYC-sen1-1* allele (Fig S.2) was constructed by tagging the *sen1-1* (kind gift from Prof Proudfoot) with a heat-sensitive degron and the MYC tag. The *sen1-1* allele confers temperature-sensitivity and is characterized by increased transcription read-through, increased presence of R-loops in the genome and associated hyper-recombination

(Mischo et al., 2011). We found that constructs not expressing the helicase domain could not complement the *td-MYC-sen1-1* allele at 37°C, in line with the established literature. By contrast fragments Sen1 (2-1901), Sen1 (2-2231), Sen1 (931-2231) and Sen1 (1095-2231) all complemented the temperature-sensitive allele upon degradation of endogenous TD-MYC-Sen1-1 proteins at 37°C and in the presence of Ubr1 (the latter is the E3 ligase required for degradation of proteins tagged with the TD degron), indicating that these constructs folded properly (Fig 4.7). As such, the inability of both the Sen1 (931-2231) and Sen1 (1095-2231) constructs to interact with replisome components is not a result of mis-folding. Taken together, these observations suggest that the first 931 residues of Sen1 are both sufficient and necessary for the interaction of the protein with the replisome.

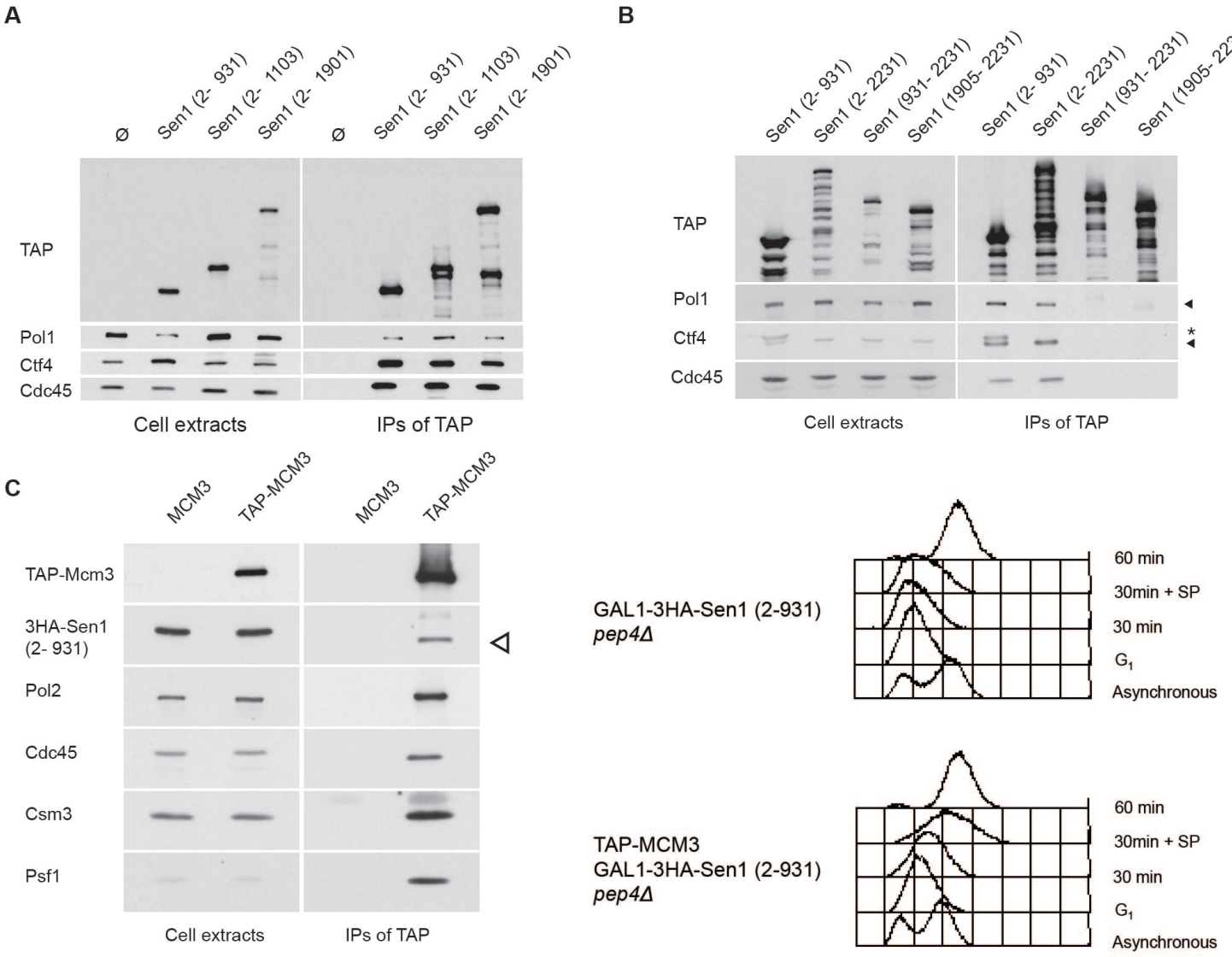

A
Ø
Sen1 (2- 931)
Sen1 (2- 1103)
Sen1 (2- 1901)
Ø
Sen1 (2- 931)
Sen1 (2- 1103)
Sen1 (2- 1901)
TAP
Pol1
Ctf4
Cdc45
Cell extracts
IPs of TAP
B
Sen1 (2- 931)
Sen1 (2- 2231)
Sen1 (931- 2231)
Sen1 (1905- 2231)
Sen1 (2- 931)
Sen1 (2- 2231)
Sen1 (931- 2231)
Sen1 (1905- 2231)
TAP
Pol1
Ctf4
Cdc45
Cell extracts
IPs of TAP
C
MCM3
TAP-MCM3
MCM3
TAP-MCM3
TAP-Mcm3
3HA-Sen1
(2- 931)
Pol2
Cdc45
Csm3
Psf1
Cell extracts
IPs of TAP
GAL1-3HA-Sen1 (2-931)
pep4Δ
60 min
30min + SP
30 min
G1
Asynchronous
TAP-MCM3
GAL1-3HA-Sen1 (2-931)
pep4Δ
60 min
30min + SP
30 min
G1
Asynchronous

Figure 4.6. Residues spanning 2-931, that include the Sen1 N-terminal domain, are responsible for interaction with the replisome. (A) Cells carrying different N-terminally tagged Sen1 fragments under the *GAL1* promoter were grown in YP-Raff at 24°C to a density of 0.7 x 10^7 cells/ml in 250 ml cultures. The cells were arrested in G_1 using α-factor to a final concentration of 7.5 μg/ml for 3 h. After successful arrest, the YPGAL medium was substituted for YP-Raff, keeping α-factor at a concentration of 7.5 μg/ml for a further 35 min. The galactose in the new medium triggers expression of the constructs. The α-factor was removing by washing with fresh YPGAL and the cells were released in S phase in YPGAL at 24°C. Cells were harvested 30 min after release. The samples were used for IPs using TAP beads. (**B)** Similar to (A) but with modifications. Cells were grown in 1 l cultures and YPGAL was used throughout the experiment. Arrows indicate the correct position of proteins blotted against (here, Pol1 and Ctf4) whilst the star symbol (*) indicates the position of the TAP-tagged protein that is recognized by the Ctf4 antibody. **(C)** Cells carrying the *GAL1-3HA-SEN1 (2-931)* allele with or without the *TAP-MCM3* allele were grown in YP-Raff at 24°C to a density of 0.7 x 10^7 cells/ml in 250 ml cultures. The cells were arrested in G_1 using α-factor to a final concentration of 7.5 μg/ml for 3 h. After successful arrest, the YPGAL medium was substituted for YP-Raff, keeping α-factor at a concentration of 7.5 μg/ml for a further 35 min. The cells were then released from G_1 arrest by washing out the α-factor with fresh YPGAL, followed by release in S phase in YPGAL at 24°C. Cells were harvested 30 min after release. FACS samples were taken for asynchronous cultures (Asyn), at G_1 arrest (G_1), after preparation of the G_1 (G_1 + SP) sample, at point of harvest in S phase (30 min), after preparation of the the S phase sample (30 min +SP) and 30 min after S phase harvest (60 min). The samples were used for IPs using TAP beads. The arrow indicates the correct position of the 3HA-Sen1 (2-931) construct. Strains used in this experiment: CS**1711**, CS**1714**, CS**1852**, CS**1933**, CS**1941**, CS**1942**, CS**1943**, CS**1957** and CS**1958**.

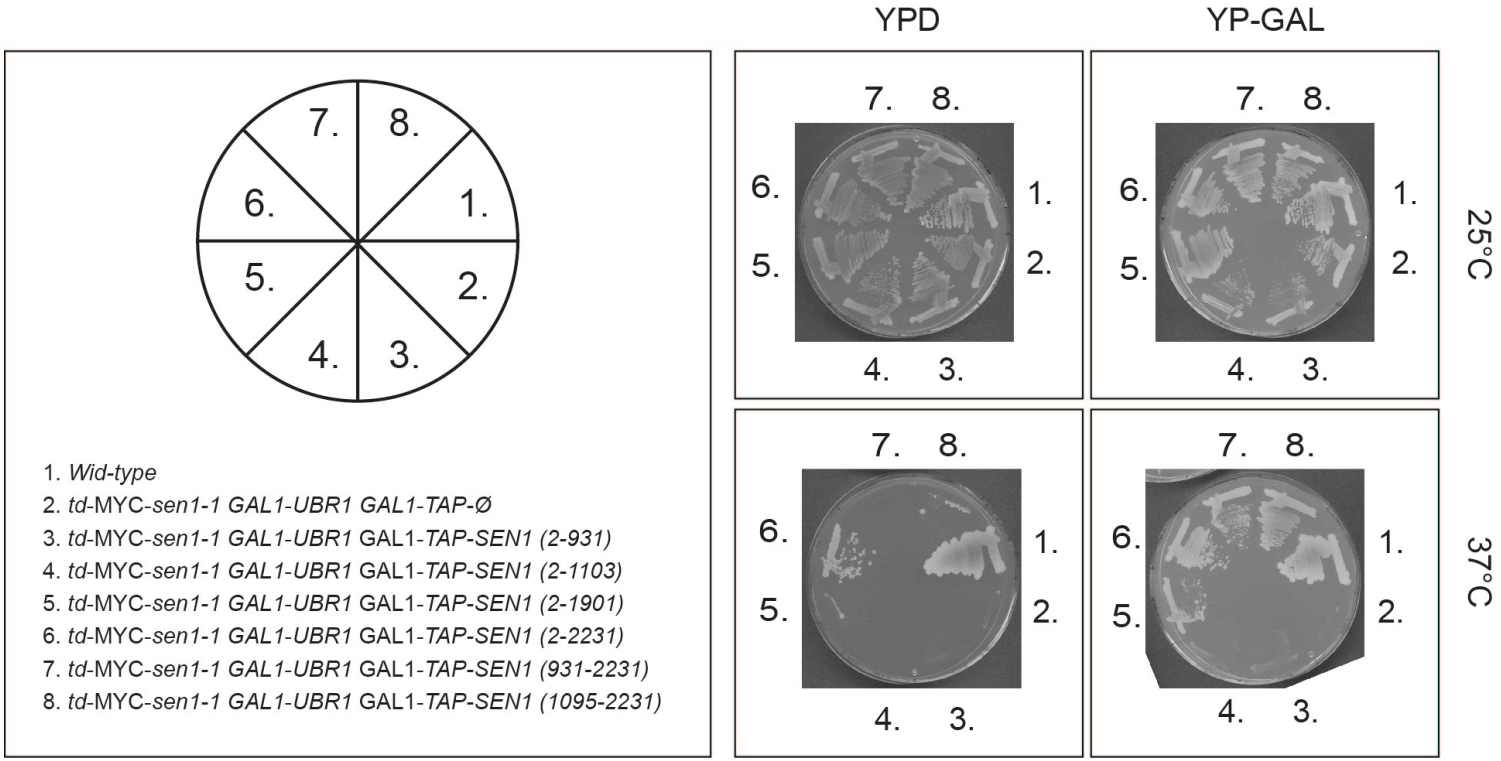


Figure 4.7. The last 1136 residues of Sen1 are sufficient to complement the temperature-sensitive of the *td-MYC-sen1-1* allele. To determine whether the different constructs of Sen1 folded correctly, they were crossed into cells carrying the *td-MYC-sen1-1* and *GAL1-UBR1* alleles. The *td-MYC-sen1-1* allele is temperature-sensitive at 37°C and induction of Ubr1 with galactose quickly leads to degradation of the TD-MYC-Sen1-1 protein by virtue of its temperature degron (TD). The strains were grown at either 24°C or 37°C, in either YPD to suppress expression of both Ubr1 and the Sen1 fragments or in YPGAL to trigger their expression. Strains used in this experiment: CS**1**, CS**2056**, CS**2058**, CS**2061**, CS**2062**, CS**2184**, CS**2188** and CS**2451**.

4.4 YEAST-TWO-HYBRID ANALYSIS REVEALS RAD5 AS A NOVEL INTERACTOR OF SEN1

Having determined that the N-terminal of Sen1 is required for interaction with the replisome, we then wanted to identify the preferred interactor or interactors through which this interaction is achieved. There are two ways to identify novel physical interactors of a protein of interest. These are the yeast-two-hybrid (Y2H) and immuno-precipitation followed by mass spectrometric analysis. In both assays, either the entire protein of interest or its fragment(s) may be used as bait.

We initially attempted to identify the preferred interactor of Sen1 by using a high-throughput, optimized variant of the yeast-two-hybrid (Y2H) assay, conducted by Hybrigenics. In brief, *SEN1 (2-931)* was cloned juxtaposed to the *GAL4* DNA binding domain. We chose this fragment and not the full-length gene based on our earlier observations that the full-length protein is more labile and less well expressed than Sen1 (2-931). Additionally, given the role of Sen1 in transcription termination, we reasoned that expressing the full length protein as a second copy (as bait) could impair global levels of proteins, including those of the candidate preys. Moreover, because of the nature of Y2H screens, the helicase domain of Sen1 could also impair expression of the reporter genes. The *SEN1 (2-931)* construct was cloned in *MATa* yeast cells that were then mated to a collection of *MATα* cells of the same background that had been previously transformed with a library of yeast (*S. cerevisiae*) genomic DNA cloned in proximity and in frame with a *GAL4* activating domain. The resulting diploids were tested for grown on synthetic medium lacking histidine. Of the 97 million interactions tested, 380 positive clones were obtained. The different positive hits were scored for non-specificity and false-negatives and the clones were sequenced to identify the prey proteins. Twelve potential interactors were identified (Table 4.1).

Table 4.1. List of potential Interactors of Sen1 Identified in the Commercial Y2H Screen (Hybrigenics)

Potential Interactors of Sen1	Function	Number of Clones
Prs2	Involved in amino acid and nucleotide biosynthesis	96
Prs3	Involved in amino acid and nucleotide biosynthesis	39
Prs4	Involved in amino acid and nucleotide biosynthesis	28
Rad2	Nucleotide excision repair	8
Rad5	Involved in DNA damage tolerance pathway	1
Msh6	Mismatch repair during mitosis and meiosis	1
Nap1	Histone chaperone	15
Lro1	Triglyceride synthesis	12
Dal81	Involved in nitrogen degradation pathways	1
Sco1	Required for cytochrome C oxidase activity and Respiration	3
Flo8	Transcription factor required for flocculation	1
Nrd1	Involved in Sen1-dependent transcription termination	23

No replisome component was identified in the Y2H screen. Although no screen can claim to be exhaustive, the failure to detect any replisome component could indicate that the preferred interactor of Sen1 within the replisome is not expressed at significantly high levels to achieve detectable levels of interaction with Sen1 in the screen. The tags used for the bait and prey proteins could also interfere with physiological interactions. Alternatively, it is possible that such an interaction is too transient to be detected by the screen. Sen1 could also interact with the replisome through a sub-complex made up of either distinct or identical proteins (such as Polymerase ε and the Ctf4-trimer respectively) in such a way that it would not be possible to recreate this interaction artificially in a Y2H screen (Simon et al., 2014). Finally, post-translational modifications of either the bait or prey may be important for interaction and the constructs used in the screen may not be modified adequately.

Of the possible interactors of Sen1 identified in the screen, Rad2 and Prs3 (a protein involved in both nucleotide and amino acid biosynthesis) had been identified previously in a Y2H screen using the Sen1 (2-975) construct as bait,

closely corresponding to the bait used here (Ursic et al., 2004). Because of the high-throughput nature of the commercial screen, we wanted to test whether Msh6, Nap1, Nrd1 and Rad5 were true interactors of Sen1. To do so, I obtained bait and prey plasmids that I transformed in a Y2H tester strain (CS114). In brief, cells were transformed simultaneously with either an empty bait plasmid or a bait plasmid expressing the *SEN1 (2-931)* construct together with either an empty prey plasmid or plasmids expressing fragments of either *MSH6*, *NAP1*, *NRD1* or *RAD5*. As an internal control, we also used prey plasmids expressing fragments of Prs2 (related to Prs3) and Rad2. Cells expressing different pairs of bait and prey plasmids were then growth on selective medium with or without 3-amino- 1,2,4-triazole (3-AT), a competitive inhibitor of the *HIS3* gene in *S. cerevisiae*. Cells expressing *NAP1* and *MSH6* constructs as prey did not grow on selective medium suggesting that the proteins they encode for do not interact physically with Sen1. On the other hand, cells expressing *SEN1 (2-931)* and *RAD5 (13-265)* grew on selective medium (Fig 4.8). Importantly, these cells grew better on selective medium compared to cells expressing both the *SEN1 (2-931)* bait and the *RAD2 (501-869)* prey. This indicates that Rad5 genuinely interacts with the first 931 residues of Sen1 and that the Rad5 (13-265) construct is sufficient for this interaction. This observation merits further investigation in the future.

The Y2H results also indicate that Sen1 and Nrd1 interact through their own respective N-termini (Fig 4.8). Previously, Sen1 was thought to only interact physically with Nab3 and not Nrd1 (Nedea et al., 2008). Given that Nab3 and Nrd1 form a hetero-dimer, these observations indicate that all three components of the NNS complex can form direct physical contact with one another.

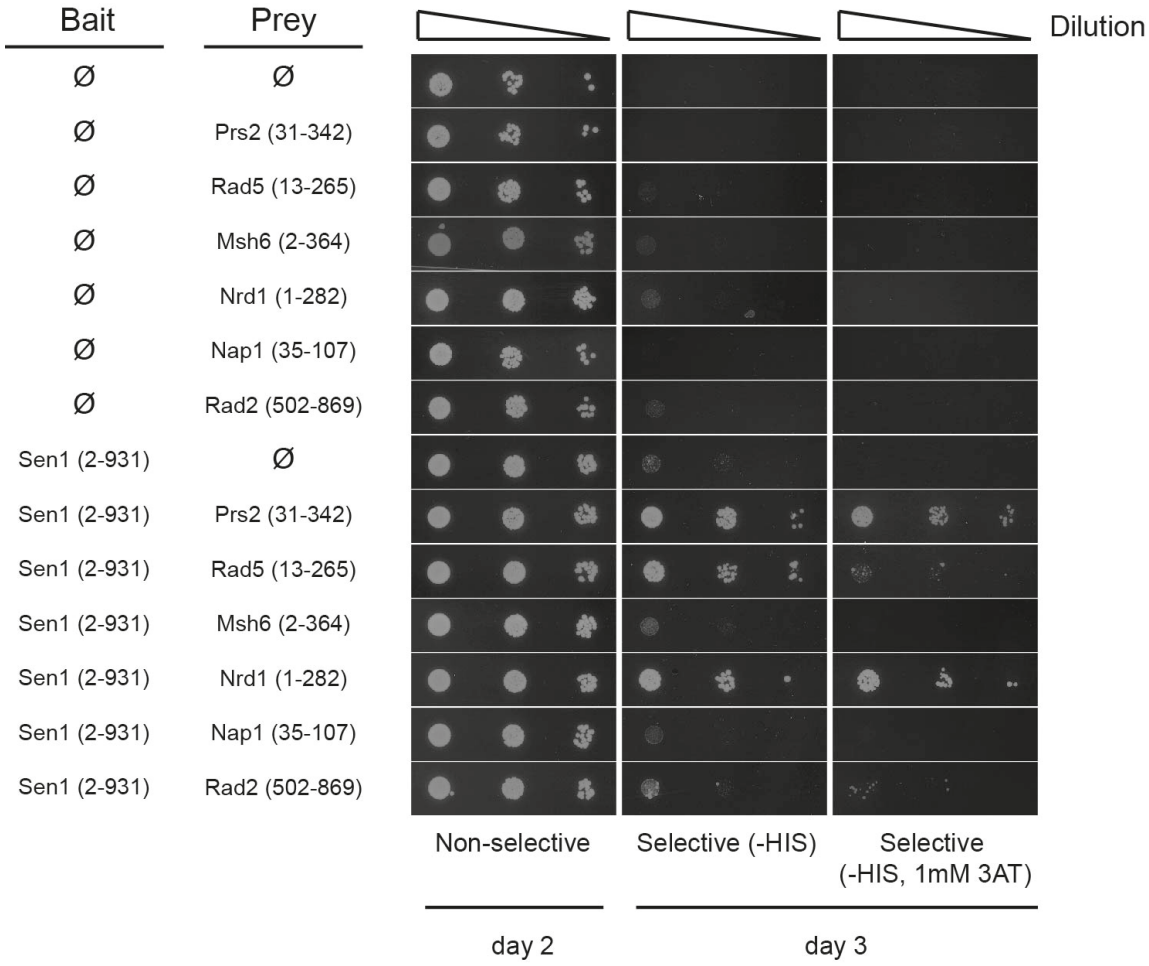


Figure 4.8. Yeast-two-hybrid screen using the GAL4-Sen1 (2-931) as bait. Diploid cells were transformed with a pair of bait and prey plasmids. The bait vector encoded either a GAL4-binding domain on its own or for the GAL4-bindind domain bound to the Sen1 (2-931) fragment. The prey plasmids encoded either for the GAL4-activation domain on its own or the activation domain bound to candidates binding partners of Sen1 (2-931). The diploids were then grown on either non-selective medium or in selective media (synthetic medium with amino acids lacking histidine with or without the histidine inhibitor, 3AT. Strain used in this experiment: CS**114**.

4.5 MASS SPECTROMETRIC ANALYSIS OF SEN1 (2-931) IPS IN DIFFERENT PHASES OF THE CELL CYCLE REVEAL EITHER CTF4 OR GINS TO BE THE INTERACTOR OF SEN1 WITHIN THE REPLISOME

As the Y2H screen did not reveal the identity of the replisome interactor(s) of Sen1, we decided to carry immuno-precipitations of Sen1 (2-931) followed by mass spectrometric analysis. We aimed to exploit the dynamic nature of replisome assembly, expecting that Sen1 (2-931) would interact with all or most of the components of the replisome in S phase but would interact uniquely with one or a few replisome components throughout the cell cycle. The latter interactors would be identified as the preferred interactor(s) of Sen1 vis-à-vis the replisome.

To obtain sufficiently concentrated protein samples for mass spectrometric analysis, four independent harvests of 7 x 10^9 cells, grown in YPGAL, and carrying the *GAL1-TAP-SEN1 (2-931)* allele were conducted for each phase of the cell cycle (G_1, S and G_2). For the control (no bait), four independent harvests of 2 x 10^{10} cycling cells, carrying the *GAL1-TAP-∅* allele, were carried out. For each biological replicate, FACS profile were run to ensure that cells were harvested in their correct respective cycles (Fig 4. 9). The samples were lysed by using a motorized pestle and mortar cooled to −80°C using liquid nitrogen and thawed to 4°C. Universal nuclease (Pierce) was added to a final concentration of 1.6U/µl and the resulting mixture was incubated at 4°C for 40 min. Thereafter, the mixture was clarified by centrifugation at 12,500 rpm in a high speed centrifuge and at 32,500 rpm in an ultra-centrifuge. TAP-Sen1 (2-931) was immuno-precipitated using IgG bound to Dynabeads in a 100 mM Hepes, 50 mM potassium acetate solution. After three washes, TAP-Sen1 (2-931) and its co-precipitates were cleaved off the Dynabeads using AcTEV protease at 24°C for 2 h. After cleavage, the resulting CBP-Sen1 (2-931) and its co-precipitates were incubated with pre-washed calmodulin beads at 4°C for 2 h and, after washing, the proteins were removed from the

calmodulin resin by boiling in 30 μl of 1 x Laemmli buffer. This simultaneously concentrates the proteins. The samples from the four biological replicates were then pooled together.

Thereafter, the samples were run for a small distance on a commercially-sourced 4-12% acrylamide gel that was then cut in thin slices and processed using a robot (MS Bioworks, USA). The protein samples were successively washed, reduced, alkylated and digested by trypsin. The reaction was then analysed using a nanoLC/MS/MS (Waters NanoAcquity HPLC/ThermoFischer Q Exactive). Of the ions identified, the fifteen most abundant were selected for tandem MS. To process the data, the spectra obtained were compared to the spectra of *in silico* trypsinized *S. cerevisiae* proteins (SGD database).

The raw data is presented in Table S1 at the very end of this thesis. There was a total of 1778 unique (non-specific) peptides identified using the control samples whilst 4343, 8066 and 5826 unique peptides were identified using the Sen1 (2-931) bait in G_1, S and G_2 respectively. This was reflected in Coomassie and silver staining of the IP samples (Fig 4.9). As a first approximation, this indicates that the mass spectrometric screen successfully identified specific interactors of CBP-Sen1 (2-931) and that more protein molecules interact with the bait in S phase than either in G_1 or G_2. Proteins for which the number of peptides identified in the control was at least 2-fold smaller compared to at least one of the experimental samples were considered specific interactors of Sen1 (2-931). The remaining proteins were considered non-specific hits.

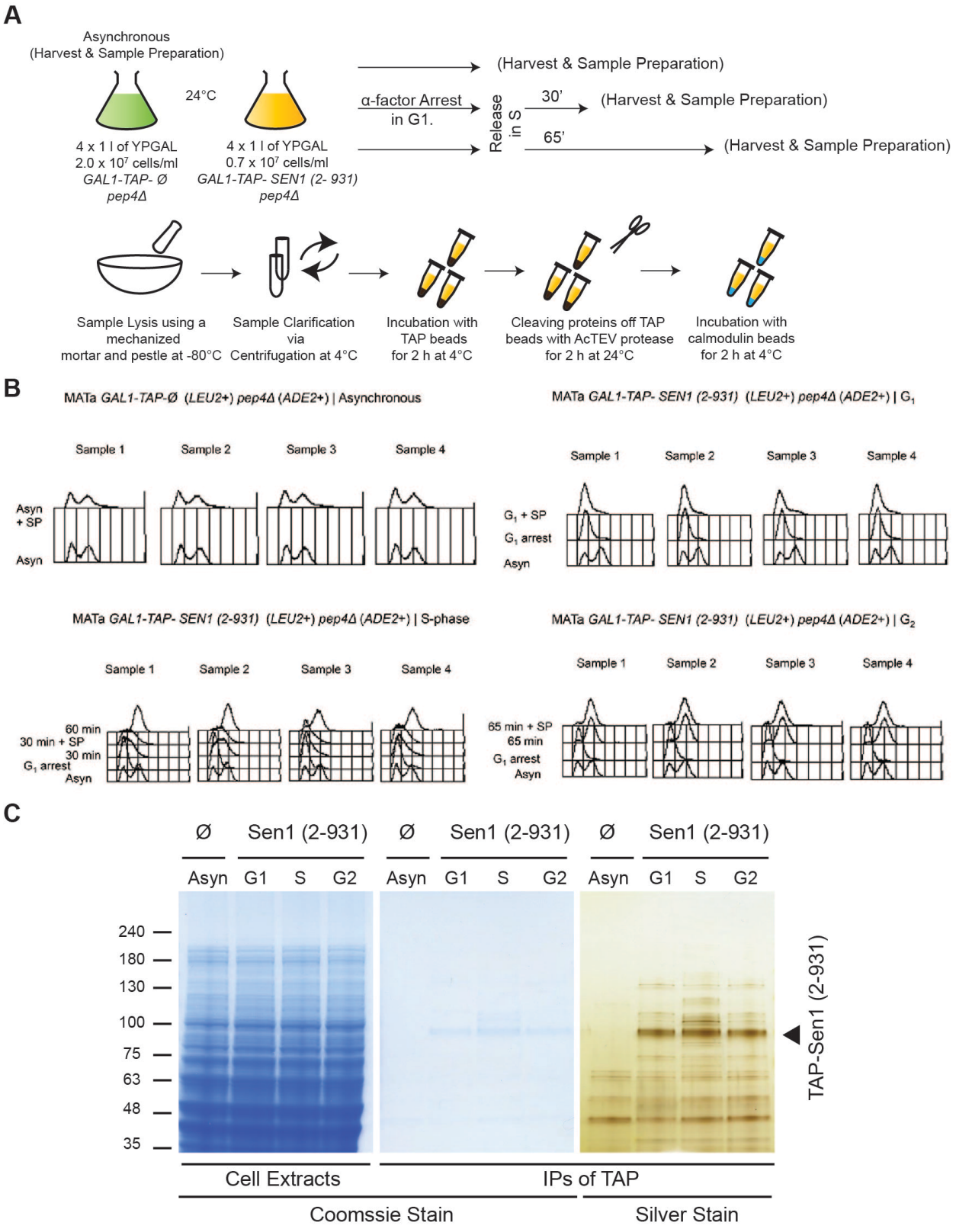


Figure 4.9. The Sen1 (2-931) fragment was used as a bait to fish for interactors using IPs followed by tandem mass spectrometry. Cells carrying the *GAL1-TAP-SEN1 (2-931)* allele were grown in YPGAL at 24°C to a density of 0.7 x 10^7 cells/ml in 4 x 1 l cultures. 4 l of samples were harvested

in either G1, S or G2 corresponding to 2.8 x 10^{10} cells for each cycle. As a control, cells carrying the *GAL1-TAP-Ø* allele were grown in YPGAL at 24°C to a density of 2 x 10^{7} cells/ml in 4 x 1 l cultures. Here, asynchronous cultures of the untagged control were used as it more cost-effective than using synchronized cultures for the untagged control. However, it should be recognized that this is not a perfect control for this experiment as the test samples consist of IPs of synchronized cultures. This can lead to bias, especially for non-specific interactors that would have fluctuating signals throughout the different phases of the cell cycle, complicating the interpretation of the results. Nonetheless, this control was adjudged to be satisfactory for the purpose of screening for interactors prior to verification by Western blotting. The total number of cells harvested for the control was 8 x 10^{10} cells. After a two-step IP, the biological replicates were pooled together and the IPs were analysed via mass spectrometric analysis. FACS samples were taken for asynchronous cultures (Asyn), at G_1 arrest (G_1), after preparation of G_1 (G_1 + SP) sample, at point of harvest in S phase (30 min), after preparation of S phase sample (30 min + SP), at point of harvest in G_2 (65 min) and after preparation of G_2 sample (65 min + SP). **(A)** Schematic of the experiment. **(B)** The FACS samples for each biological replicate was run and this shows that the cells were harvested in their correct respective cycles. **(C)** A small quantity of the IPs from the experiment were run on polyacrylamide gels and the bands visualized using Coomassie or Silver stains. Strains used in this experiment: CS**1852** and CS**1957**.

We observed that several replisome components co-precipitate with CBP-Sen1 (2-931) in S phase, including with all the components of the fork protection complex (Mrc1, Tof1 and Csm3), all components of the Polymerase α holoenzyme (Pol1, Pol12, Pri1 and Pri2), two components of the Polymerase ε holoenzyme (Pol2 and Dpb2), the topoisomerase Top1, the Ctf4 interaction hub (Villa *et al.* 2016), and all 11 subunits of the CMG helicase (Cdc45, Mcm2, Mcm3, Mcm4, Mcm5, Mcm6, Mcm7, Sld5, Psf1, Psf2, Psf3). The inability to detect two proteins within the Polymerase ε holoenzyme (Dpb3 and Dpb4) could be reflective of technical limitations of the experiments to detect ions from smaller proteins. The screen also revealed the presence of Pob3 and Spt16 (members of the FACT complex that acts as a histone chaperone). Peptides of the F-box protein Dia2 and the cullin protein Cdc53 that are required for replisome disassembly (Maric et al., 2014) were also detected in S phase IPs of CBP-Sen1 (2-931). Histone proteins were also detected (HTA2, HTB2, HHF1) (Fig 4.9). Taken together, this indicates that several sub-complexes of the replisome can co-purify with CBP-Sen1 (2-931) in S phase (Fig 4.10).

By contrast, few peptides corresponding to replisome proteins were detected in the screen for the G_1 phase, except for Ctf4 and GINS (Sld5, Psf1, Psf2 and Psf3). The signal for the GINS complex in G_1 is more convincing when normalizing for the molecular weight and number of tryptic sites within the proteins (Fig 4.11). Sen1 (2-931) then could potentially interact directly with both Ctf4 and GINS. However, given that Ctf4 is known to interact directly with Sld5 (Gambus et al., 2009, Tanaka et al., 2009a), another plausible scenario is that either Ctf4 or GINS interact directly with Sen1 in G_1. Interaction with GINS and Ctf4 could also be seen in the G_2 screen although peptides from other components of the replisome could also be seen. Since replisomes disassemble at the onset of G_2, this could be indicative of some late forks still going through the replication process.

Of interest, earlier experiments indicated that Sen1-TAP did not co-purify with replisome components outside of S phase (Fig 4.2 and 4.3). These apparently

contradictory findings can be reconciled if the last 1300 residues of Sen1 were to regulate the interaction of the protein to replisome components, perhaps through some self-inhibitory mechanism. Thus, had the full-length Sen1 been used as bait instead, the interaction with Ctf4 and GINS in G_1 might have gone undetected.

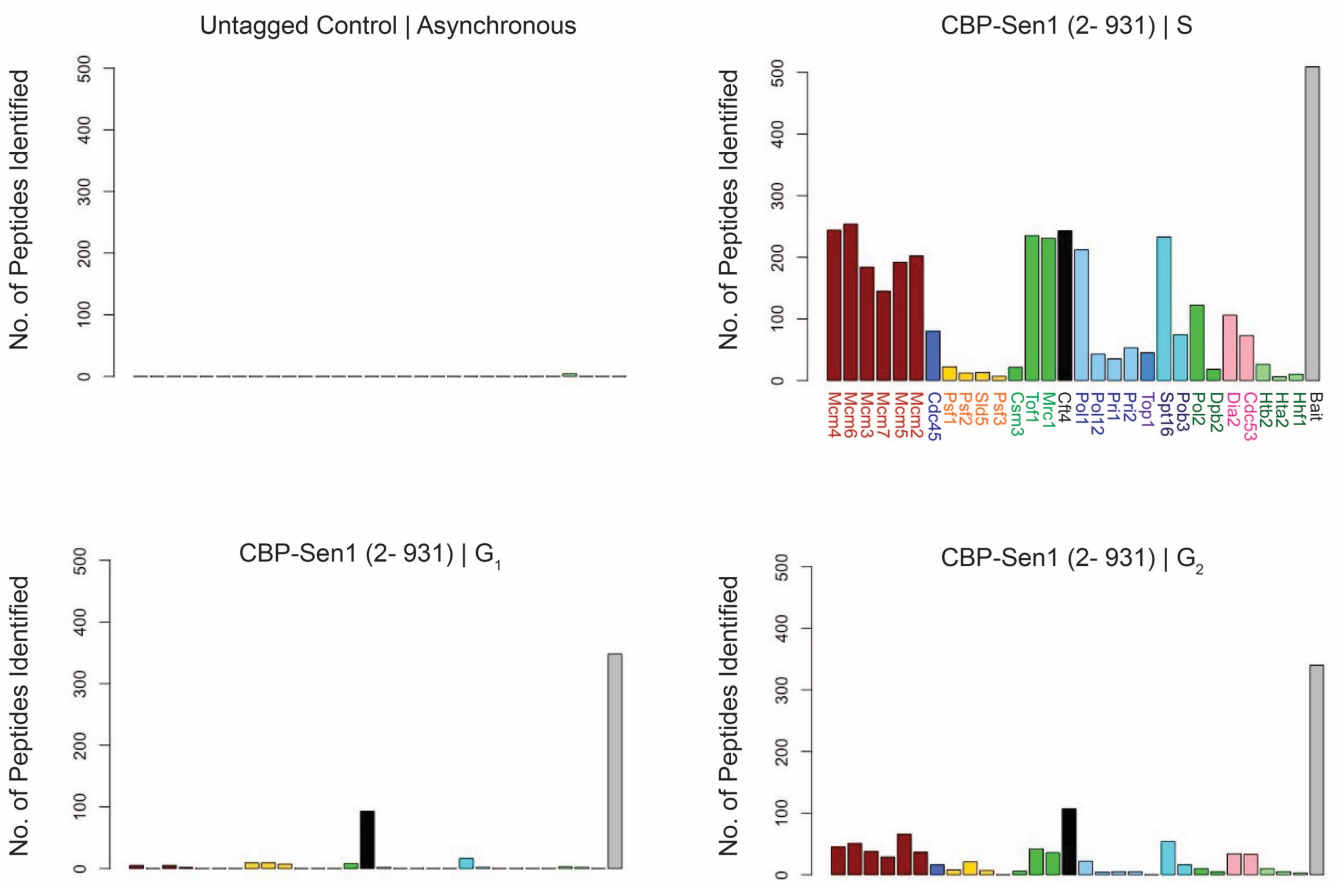


Figure 4.10. The mass spectrometric screen reveals that Ctf4, can interact specifically with Sen1 (2-931) throughout the cell cycle. Here, the raw data of the number of peptides from replisome components that were identified in the mass spectrometric screen is represented graphically. The identity of the different proteins is indicated for the S phase plot. For simplicity, the corresponding identifiers for the control, G_1 and G_2 plots are masked.

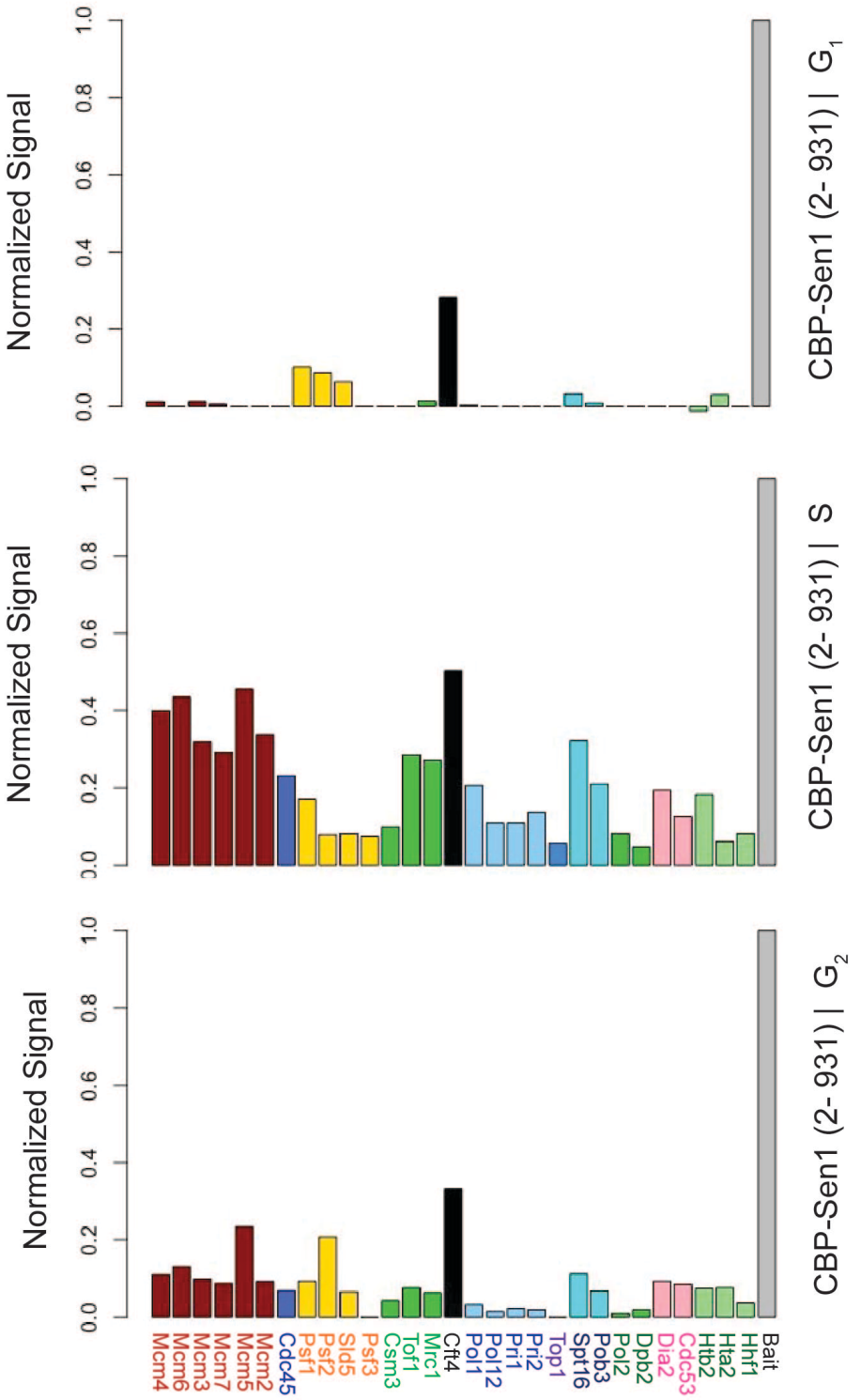


Figure 4.11. Normalizing the data from the mass spectrometric screen reveals both Ctf4 and GINS as likely interactors of Sen1. The data from the mass spectrometric screen was normalized (Section 2.5.2), taking in account the number of likely number of tryptic fragments and the molecular weights of the different proteins identified.

4.6 Sen1 (2-931) co-purifies with components of RNA Pol I, II and III throughout the cell cycle

In addition to replisome components, other proteins were identified as *bona fide* interactors of Sen1 (2-931). This includes several subunits of RNA Pol I, II and III (Table 4.2, Fig 4.12 and 4.13). Whilst subunits of RNA polymerases have previously been shown to co-purify with full-length Sen1 (Yüce and West, 2013), here we show that Sen1 (2-931) construct interacts with RNA polymerases throughout the cell cycle. It would be interesting to determine whether this also holds true for the full-length Sen1 protein and how this affects its function. Another addition to the scholarship is the way through which Sen1 interacts with RNA Pol I and III. It has previously been shown that Sen1 interacts with the CTD terminal of RNA Pol II by virtue of its N-terminal domain (Chinchilla et al., 2012). In fact, the Sen1R302W protein is characterized by loss of interaction with RNA Pol II. However, it was previously unknown through which domain Sen1 interacted with either RNA Pol I or III. Here, mass spectrometric data indicate that the first 931 residues of Sen1 are sufficient for interaction with several subunits of both RNA Pol I and III (Table 4.2, Fig 4.12 and Fig 4.13).

The peptides identified included those from the largest subunits of all three RNA polymerases (Rpa190, Rpb1 and Rpc160). Peptides from the second largest subunit was observed for both RNA Pol II (Rpb2) and RNA Pol III (Rpc128) whilst the third largest subunit common to RNA Pol I and III (Rpc40) was detected. Subunits from the stalk sub-complex were only detected for RNA Pol III (Rpc17 and Rpc25). Given that RNA Pol I, II and III share a common ancestor, it is possible that the N-terminal domain of Sen1 interacts with some common features between the RNA polymerases. However, we cannot rule out that the 1400 other residues of Sen1 have no role in interaction with the RNA polymerases.

It is also interesting to note that more peptides from RNA Pol III subunits were detected compared to RNA Pol I and II (Fig 4.13). This possibly indicates that

although Sen1 interacts with all three RNA polymerases, RNA Pol III might have the highest affinity for the N-terminal domain of Sen1. Our results could also indicate that some of the defects reported in strains carrying the *sen1-2* allele [Sen1 (976-2231)] could be a result of loss of interaction with the RNA polymerases (Sariki et al., 2016).

Table 4.2. Subunits of the RNA polymerases in *S. cerevisiae*. RNA polymerases are made up of four to six sub-complexes and each sub-complex is made up of a number of subunits. Some of these subunits are shared across all three RNA polymerases. The architecture of the different subunits within the different RNA polymerases is depicted in Figure 4.12. The numbers within the brackets are unique identifiers for each subunit.

Structures		**RNA Polymerase**		
		Pol I	Pol II	Pol III
Sub-structures				
Core	Subassembly including the largest subunit	Rpa190 (190)	Rpo21/Rpb1 (1)	Rpo31/Rpc160 (160)
		Rpb5/ Abc27 (5)	Rpb5/ Abc27 (5)	Rpb5/ Abc27 (5)
		Rpo26/Rpb6/ Abc23 (6)	Rpo26/Rpb6/ Abc23 (6)	Rpo26/Rpb6/ Abc23 (6)
		Rpb8/ Abc14.5 (8)	Rpb8/ Abc14.5 (8)	Rpb8/ Abc14.5 (8)
	Subassembly including the second largest subunit	Rpa135 (135)	Rpb2 (2)	Ret1/ Rpc128 (128)
		Rpa12/ A12.2 (12.2)	Rpb9 (9)	Rpc11 (C11)
	Third subassembly	Rpc40 (40)	Rpb3 (3)	Rpc40 (40)
		Rcp19 (19)	Rpb11 (B11)	Rcp19 (19)
		Rpb10/ Abc10 (10)	Rpb10/ Abc10 (10)	Rpb10/ Abc10 (10)
		Rpc10/ Rpb12/Abc10 (12)	Rpc10/ Rpb12/Abc10 (12)	Rpc10/ Rpb12/Abc10 (12)
Stalk		Rpa14 (14)	Rpb4 (4)	Rpc17 (17)
		Rpa43 (43)	Rpb7 (7)	Rpc25 (25)
TFIIF-like subcomplex		Rpa49	Not applicable	Rpc37 (37)
		Rpa34/ A34.5 (34.5)		Rpc53 (53)
TFIIE-related subcomplex		Not applicable	Not applicable	Rpc82 (82)
				Rpc34 (34)
				Rpc31 (31)

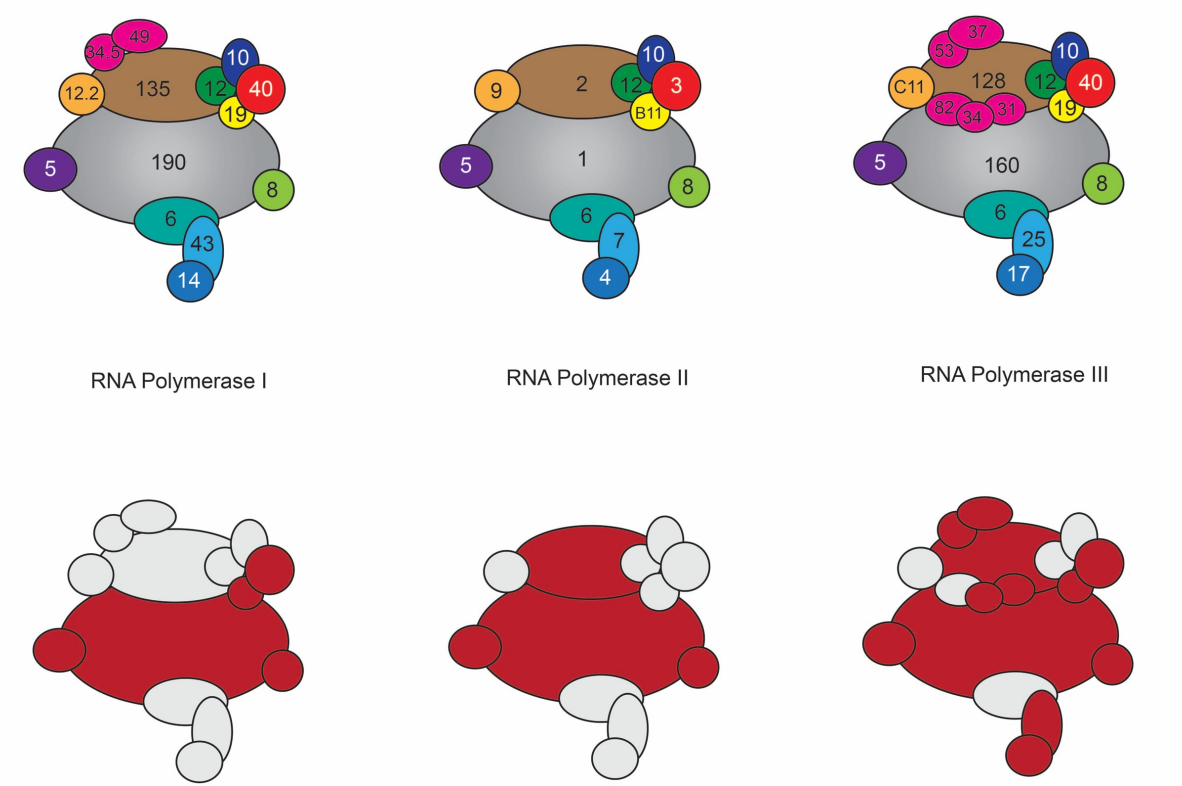


Figure 4.12.The subunits of RNA Polymerase I, II and III. Upper panel: Eukaryotic RNA polymerases are made up of four to six sub-complexes. Here, a schematic of the architecture of the three RNA polymerases from yeast are shown. The identity of the different subunits is indicated by unique numerical identifiers (Table 4.2). Lower panel: The different subunits that have been identified in the mass spectrometric screen as interactors of Sen1 (2-931) are depicted in red. The subunits that were not detected are shown in light grey. Adapted from (Wild and Cramer, 2012).

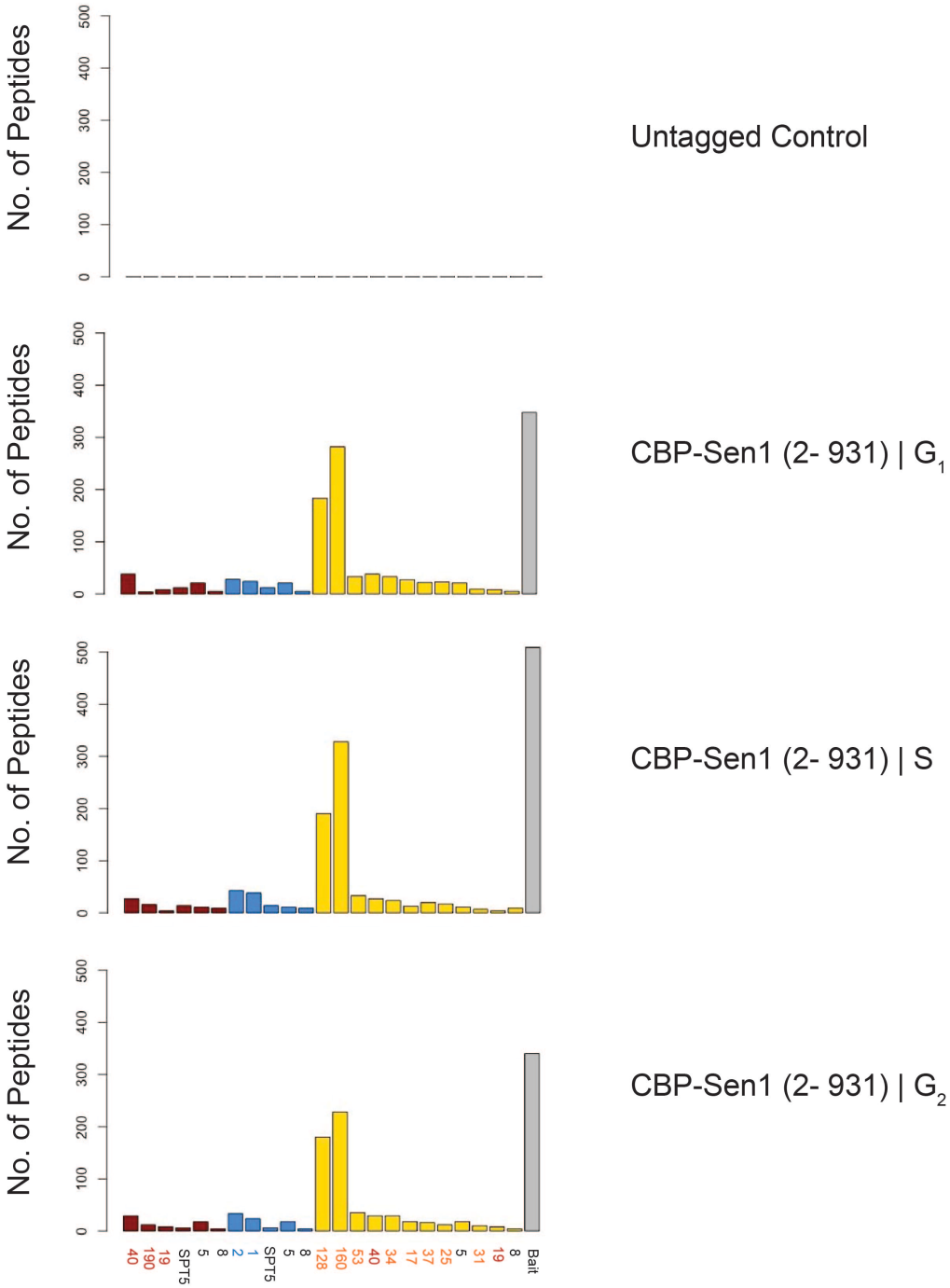


Figure 4.13. The mass spectrometric screen reveals that Sen1 (2-931) can bind to several subunits of the RNA polymerases. Here, the raw data of the number of peptides from subunits of the RNA polymerases identified in a mass spectrometric screen using Sen1 (2-931) as bait is depicted. The identity of the different subunits is indicated by unique numerical identifiers (Table 4.2). The plot also shows peptides from the Spt5 protein.

4.7 SEVERAL *TY1 COPIA-LIKE* RETROTRANSPOSONS CO-PURIFY WITH SEN1 (2-931)

Transposons (or transposable elements, TEs) are mobile genetic sequences within the genome. Class I TEs (retrotransposons) encode for a reverse transcriptase in their '*pol'* gene that they use to synthetize their cDNA anew from their own RNA transcripts. They can insert the new copies of DNA at some distance to their previous locus. This allows for amplification of the gene in the wider genome. Retrotransposons can be classified as long terminal repeat (LTR) retrotransposons, non-LTR retrotransposons and endogenous retroviruses. LTR-retrotransposons can be sub-divided into several groups, including the *Ty1-copia* retrotransposons (these include Ty1 and Ty2 retrotransposons that differ by their '*gag*' genes that code for their virus-like particle (VLP)) (Fig 4.14). It had previously been reported that R-loops associate to *Ty1* retrotransposons sites at low incidences in wildtype cells but this association is enriched in the absence of the RNase H enzymes and the topoisomerase, Top1 (El Hage et al., 2014). R-loops then are an intermediate of the *Ty1* life cycle and their targeted removal by the RNase H enzymes and Top1 would seem to silence this life cycle. Results here indicate that Sen1 interacts specifically with the **protein products** of several *Ty1* and *Ty2* retrotransposons throughout the cell cycle (Fig 4.15). How this observation would fit with the broader literature is not obvious. We use a universal nuclease in our IPs and this would suggest that identification of the retrotransposons proteins was not an artefact of protein-nucleic acid interaction but genuine protein-protein interactions. Thus, our results indicate a possible role of Sen1 in retrotransposon biology, possibly distinct from those of the RNase H enzymes and Top1.

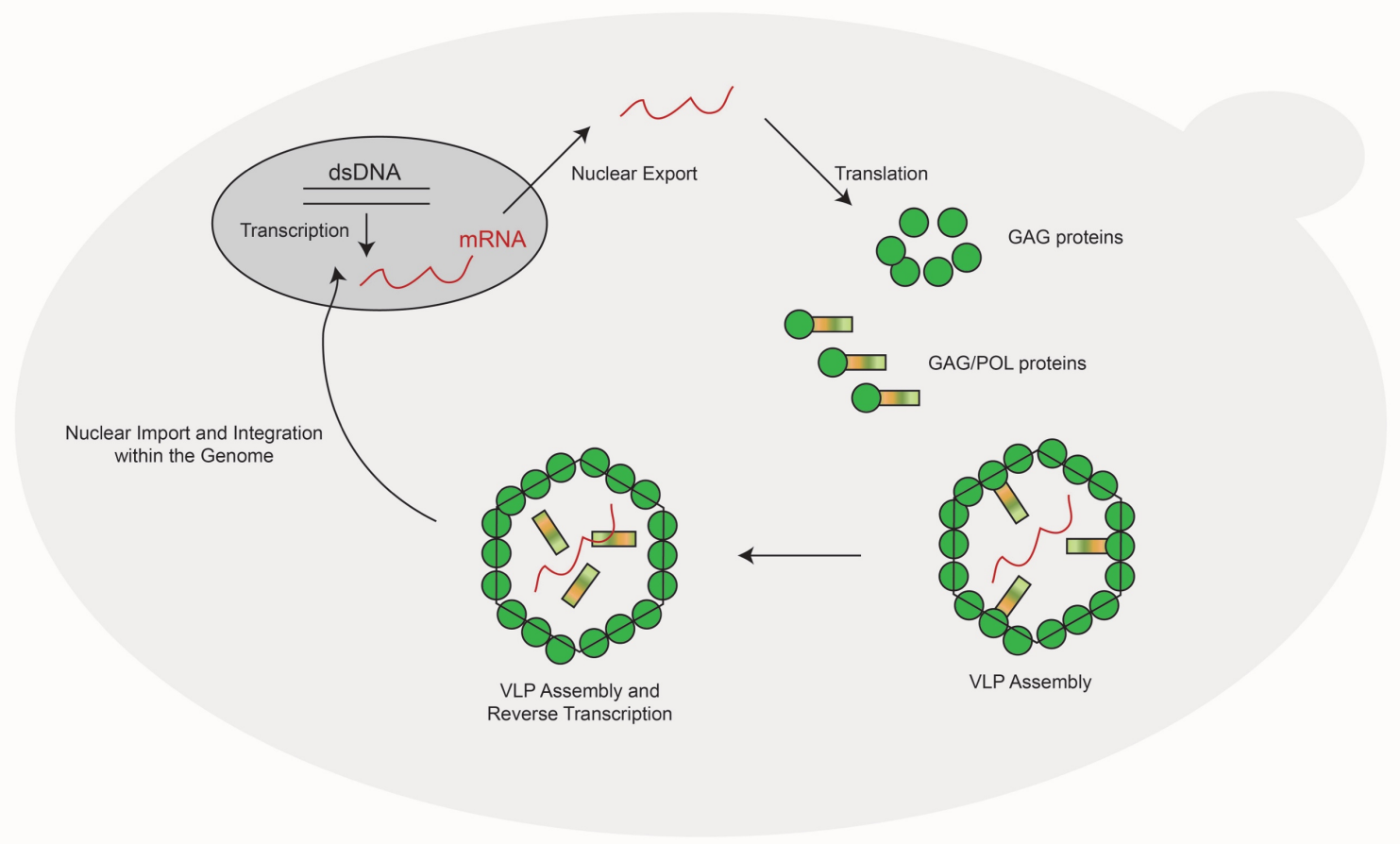


Figure 4.14. Schematic of the life cycle of retrotransposons, including *Ty1* retrotransposons.

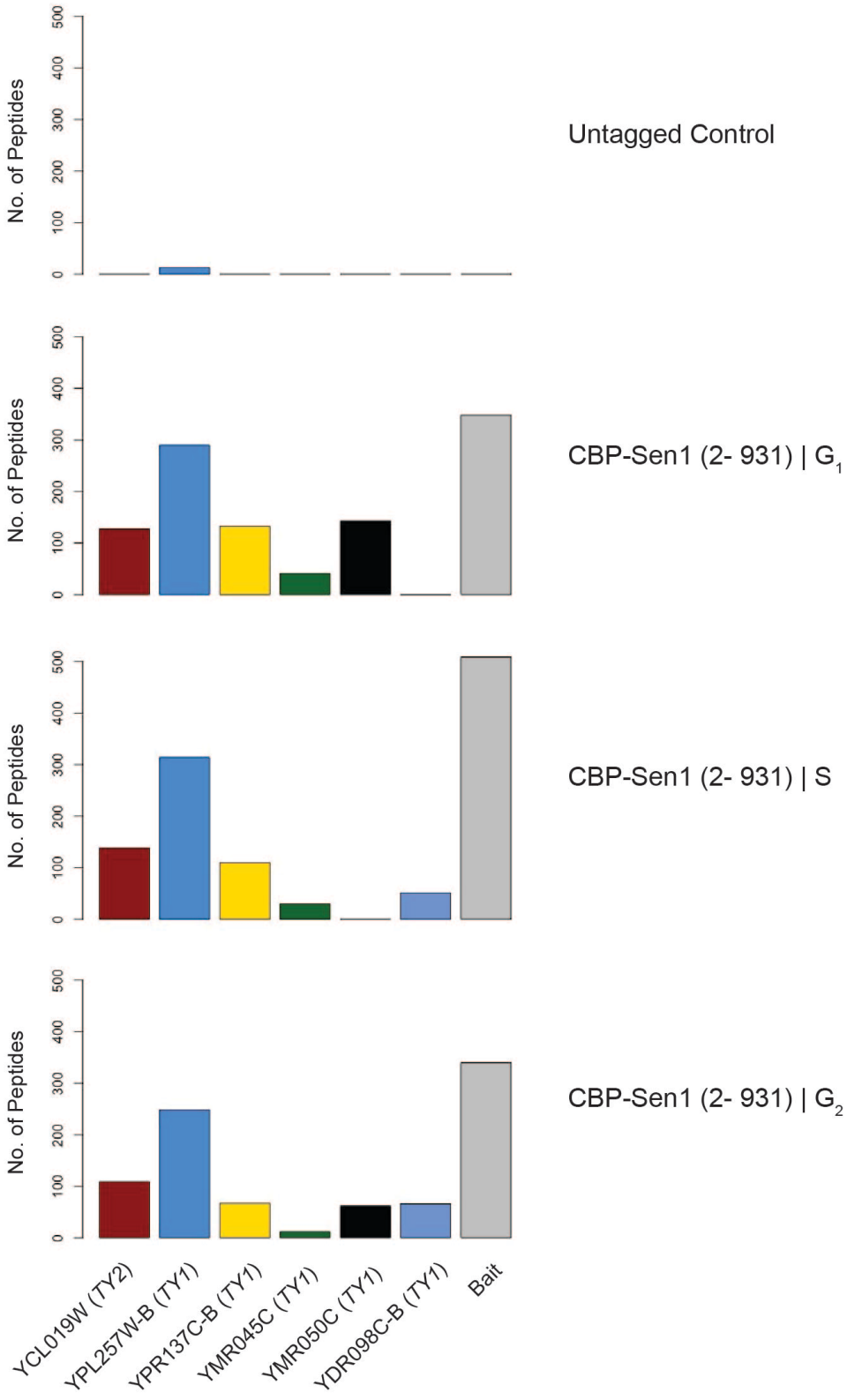


Figure 4.15. The mass spectrometric screen reveals that Sen1 (2-931) can bind to *Ty1* retrotransposons. Here, we depict the raw data of the number of peptides from *Ty* retrotransposons that were identified in the mass spectrometric screen using Sen1 (2-931) as bait.

4.8 SEN1 (2-931) REQUIRES CTF4 TO BIND TO THE GINS COMPLEX IN G_1 BUT RETAINS INTERACTION WITH REPLISOME COMPONENTS IN THE ABSENCE OF CTF4

Our mass spectrometric screen of IPs of Sen1 (2-931) revealed that the construct interacted with several components of the replisome in S phase but it interacted with both Ctf4 and GINS throughout the cell cycle. This contrasts with results of IPs of full length Sen1 where neither Ctf4 nor GINS co-precipitate with Sen1 outside of S phase. To validate the results from the mass spectrometric screen, we performed IPs of the Sen1 (2-931) construct in both G_1 and S phase and analysed their co-precipitates by Western blotting. Using this approach, we confirmed that Sen1 (2-931) can interact with both Ctf4 and GINS outside of S phase (Fig 4.16). This suggests that Sen1 (2-931) interacts with other components of the replisome either through one of Ctf4 or GINS alone, or through both of them. To differentiate between these two possibilities, we carried out IPs of TAP-Sen1 (2-931) in the absence of Ctf4 in both G_1 and S phase. We found that, in cells carrying the *ctf4Δ* mutation, no replisome components (including GINS) co-precipitated with Sen1 (2-931) in G_1 (Fig 4.16). This strongly suggested that Sen1 (2-931) interacts with the replisome through Ctf4 and not GINS.

However, whilst *ctf4Δ* did lead to a substantial reduction in the levels of replisome components co-precipitating with Sen1 (2-931) in S phase, it did not lead to a complete loss of interaction between Sen1 (2-931) and the rest of the replisome. In fact, the S phase results from the *ctf4Δ* mutant are difficult to interpret as these cells are characterized by fewer forks, naturally leading to reduced numbers of replisomes (Tanaka et al., 2009a), potentially hampering the read-out of the IPs in S phase. Taken together however, these results could be indicative of two models. Either Ctf4 is one of at least two replisome interactors of Sen1 where the other putative interactors of Sen1 associate with Sen1 (2-931) solely in S phase. Or the reduction of Sen1 in S phase GINS IPs in the absence of *CTF4* simply correlates with fewer replisomes in *ctf4Δ* cells.

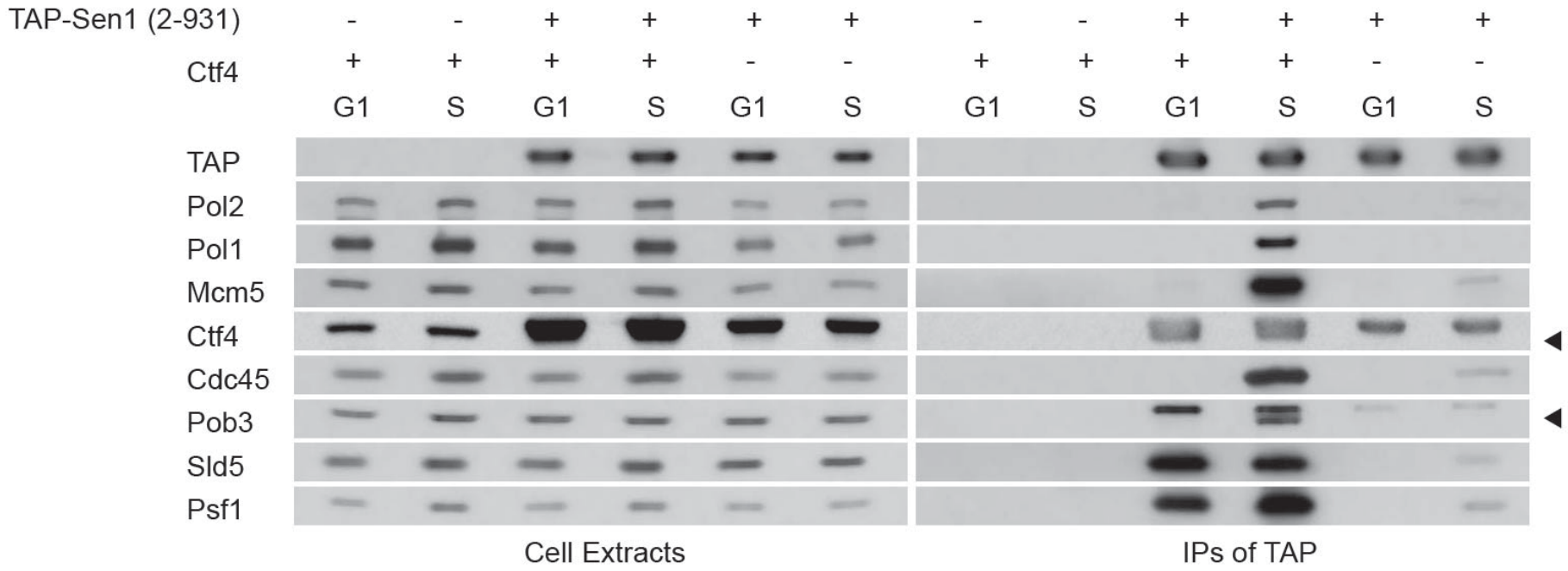


Figure 4.16. Sen1 (2-931) interacts with both Ctf4 and GINS throughout the cell cycle but Sen1 (2-931) cannot interact with GINS in G_1 in the absence of Ctf4. Cells carrying the *GAL1-TAP-SEN1 (2-931)* allele with or without *ctf4Δ* were grown in YPGAL to a density of 0.7 x 10^7 cells/ml in 250 ml cultures at 30°C. The cells were arrested in G_1 using α-factor to a final concentration of 7.5 μg/ml for 3 h. After successful arrest, the cells were either harvested and sample-prepared or were released from α-factor arrest by washing twice in fresh YPGAL, re-suspended in 250 ml YPGAL and released in S phase at 30°C. The cells were harvested 20 min post-release. The samples were then used for IPs using TAP beads. Strains used in this experiment: CS**1852**, CS**1957** and CS**2603**. For the panel corresponding to the Ctf4 signal, two bands are visible. The slower-running band corresponds to the TAP-Sen1 (2-931) construct (~128 kDa) whilst the faster-running protein corresponds to Ctf4 (~104 kDa). For the panel corresponding to the Pob3

signal, two bands are also visible. The slower-running signal is an unknown protein. The latter is unlikely to be a non-specific interactor as it is not present in IPs of the untagged control and, in the absence of Ctf4, its retention in Sen1 (2-931) IPs is reduced in both G_1 and S phase. The faster running protein corresponds to Pob3 (~63 kDa). For both Ctf4 and Pob3 panels, the arrows indicate the position of the specific signals for these two proteins.

4.9 CTF4 INTERACTS WITH SEN1 BY VIRTUE OF ITS C-TERMINAL DOMAINS

Ctf4 is made up of a WD40 domain in its N-terminus, a central β-propeller domain and an α-helical structure at its C-terminus. A second WD40 was also found in its C-terminal region (Simon et al., 2014). Replisome-associated proteins that bind to Ctf4 do so at the C-terminal WD40 by virtue of CIP boxes. We wondered whether Sen1 also binds to this domain. We carried out IPs of the Sen1 (2-931) construct in G_1 encoding either Ctf4-9MYC, 9MYC-Ctf4 or N- or C-terminally MYC-tagged truncated variants of Ctf4. In the absence of TAP-Sen1 (2-931) protein, no MYC-tagged proteins were detected in the IPs. On the other hand, N- and C-terminally MYC tagged Ctf4 co-precipitated with the Sen1 construct with identical intensities suggesting that tagging of Ctf4 at either of its termini does not affect its interaction with the Sen1 construct (Fig 4.17). The 9MYC-Ctf4 (351-927) construct also interacted with TAP-Sen1 (2-931) indistinguishably from both 9MYC-Ctf4 and Ctf4-9MYC, suggesting that the N-terminal WD40 is not required for this interaction.

Absence of the last 544 residues of Ctf4, that corresponds to loss of both the central β-propeller domain and the C-terminal α-helical, completely abrogated its interaction with the Sen1 construct in G_1 whilst truncation of the last 86 residues led to a marked reduction but not a complete loss of this interaction. This result suggests that Ctf4 interacts with Sen1 mainly by virtue of its C-terminal WD40 domain, but even when this domain is ablated, Ctf4 can still bind weakly to Sen1.

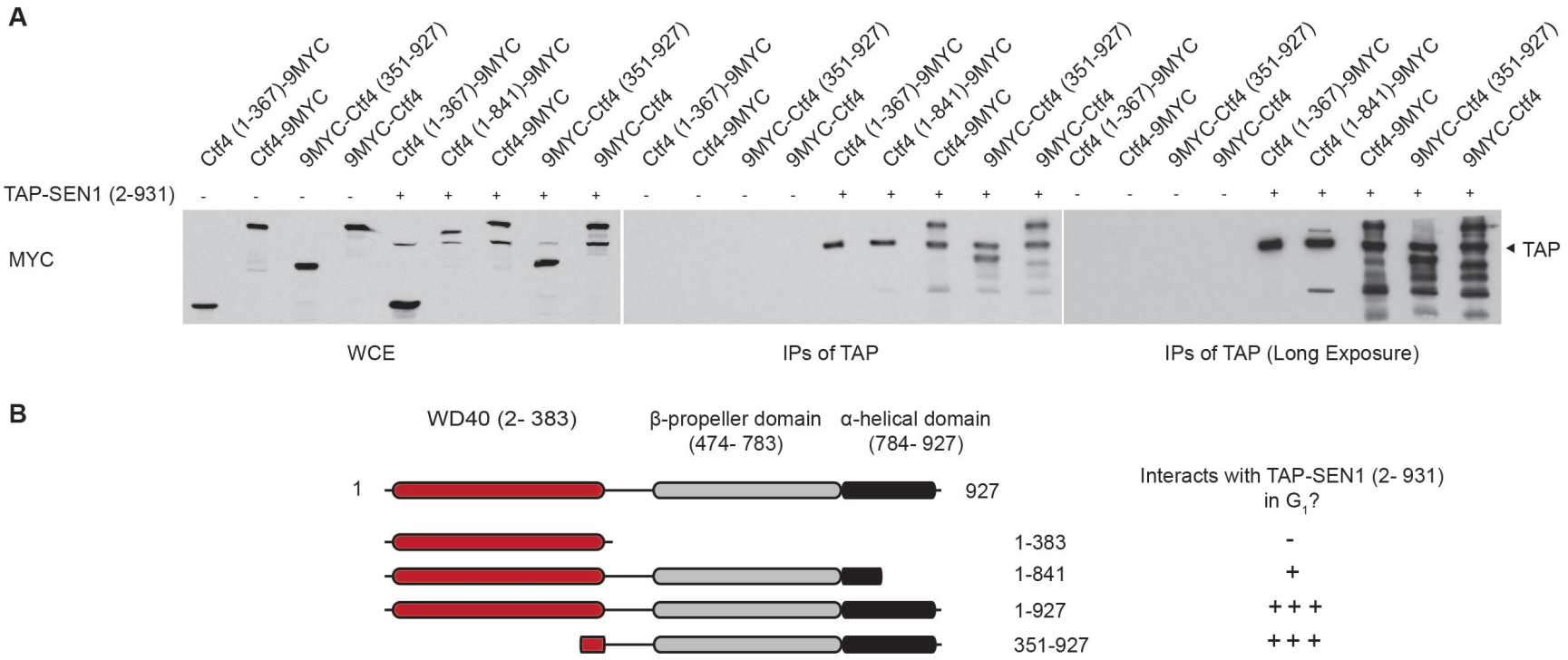


Figure 4.17. Sen1 interacts with Ctf4 by virtue of the C-terminal domains within Ctf4. Isogenic yeast cells encoding MYC-tagged alleles of either full-length or truncated *CTF4* with or without the *GAL1-TAP-SEN1 (2-931)* allele were grown in YPGAL at 24°C to a density of 0.7 x 10^7 cells/ml. The cells were synchronized in G_1 by addition of α-factor to a final concentration of 7.5 µg/ml for 3 h. The cells were harvested at −80°C and ground using a motorized pestle and mortar at −80°C. The liberated proteins were immuno-precipitated using magnetic beads conjugated to IgG antibodies. **(A)** Whole cell extracts and IP samples immunoblotted with antibodies that recognize MYC-tagged proteins. **(B)** Schematic summarising the constructs of Ctf4 that can bind to Sen1 (2-931) in G_1. Strains used in this experiment: CS**2798**, CS**2799**, CS**2800**, CS**2801**, CS**2802**, CS**2803**, CS**2804**, CS**2805** and CS**2806.**

To determine the importance of Ctf4 on Sen1's interaction with the replisome, we carried out IPs of GINS in S phase with or without Ctf4. As a control, we also carried out GINS IPs in the absence of other non-essential components of the replisome, including Top1 and subunits of the fork protection complex (Csm3, Mrc1 or Tof1). We saw reduced retention of Sen1 in GINS IPs in the absence of Ctf4 but not with in the absence of Top1 or of individual subunits of the FPC. In the absence of Ctf4, the GINS IP also retained fewer components of other replisome components as expected given that *ctf4Δ* leads to a reduction in the number of active replication forks per unit mass (Fig 4.18). This suggests that Sen1, like Sen1 (2-931) interacts with Ctf4 but that either this interaction is redundant for Sen1's interaction with the rest of the replication machinery or that, Sen1 can somehow compensate for the absence of Ctf4 by binding to some other protein(s).

We then decided to exploit the fact that different sub-complexes of the replisome interact with one another with different affinities. We investigated the retention of full length Sen1 in S phase IPs of the MCM2-7 complex (TAP-Mcm3), Polymerase α (Pol12-TAP) and Polymerase ε (Dpb2-TAP). We reasoned that should Ctf4 be the sole interactor of Sen1, the amount of Sen1 co-precipitating with replisome components should be proportional to the amount of Ctf4 also co-precipitated. Interestingly, this particular ratio was not constant, suggesting that Ctf4 is not the sole replisome interactor of Sen1. Interestingly, the amount of Sen1 co-precipitating was higher in IPs of Polymerase α and ε, compared to IPs of MCM2-7 (effectively IPs of the CMG) (Fig 4.19).

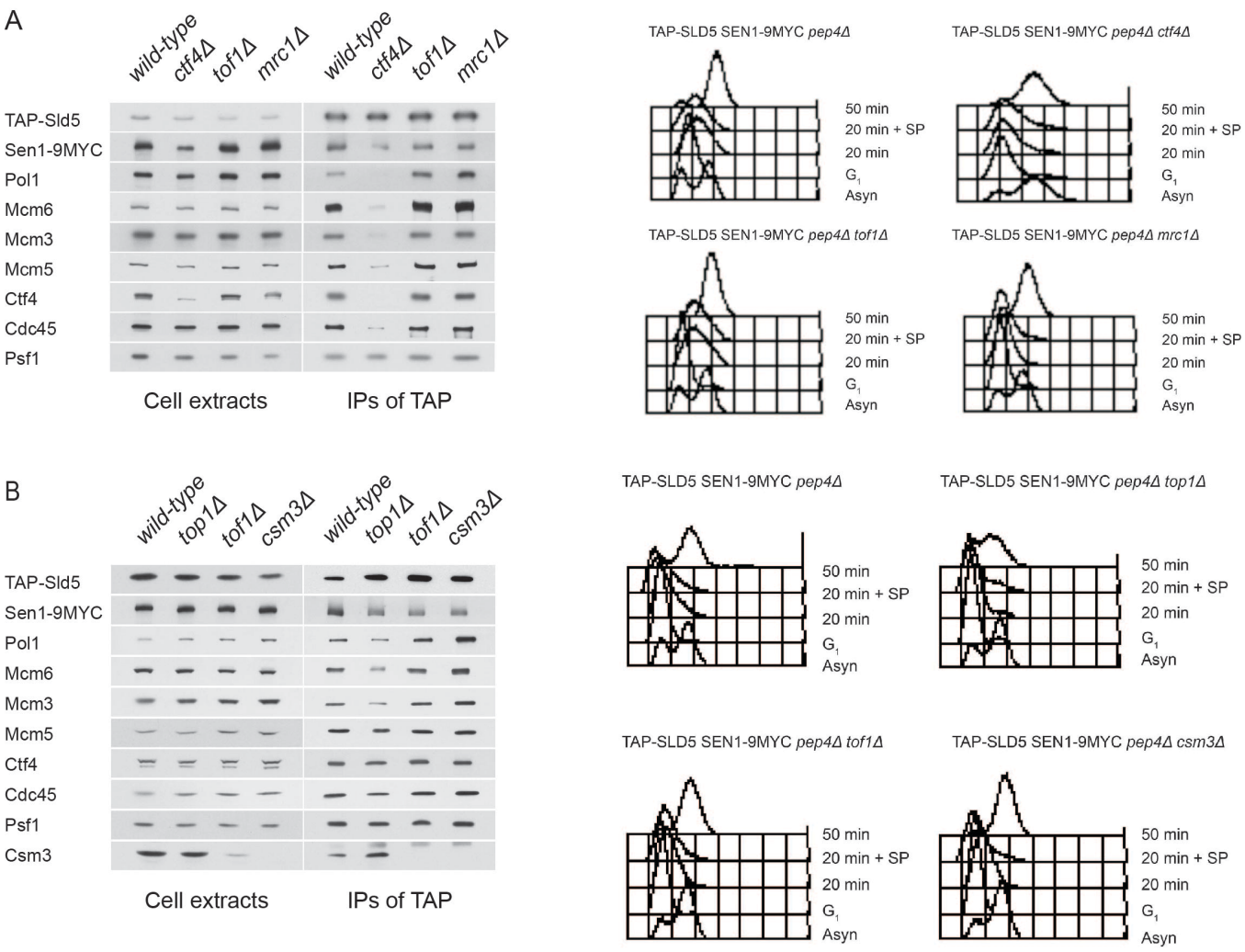

A
wild-type
ctf4Δ
tof1Δ
mrc1Δ
wild-type
ctf4Δ
tof1Δ
mrc1Δ
TAP-Sld5
Sen1-9MYC
Pol1
Mcm6
Mcm3
Mcm5
Ctf4
Cdc45
Psf1
Cell extracts
IPs of TAP
TAP-SLD5 SEN1-9MYC pep4Δ
TAP-SLD5 SEN1-9MYC pep4Δ ctf4Δ
TAP-SLD5 SEN1-9MYC pep4Δ tof1Δ
TAP-SLD5 SEN1-9MYC pep4Δ mrc1Δ
50 min
20 min + SP
20 min
G1
Asyn
B
wild-type
top1Δ
tof1Δ
csm3Δ
wild-type
top1Δ
tof1Δ
csm3Δ
TAP-Sld5
Sen1-9MYC
Pol1
Mcm6
Mcm3
Mcm5
Ctf4
Cdc45
Psf1
Csm3
Cell extracts
IPs of TAP
TAP-SLD5 SEN1-9MYC pep4Δ
TAP-SLD5 SEN1-9MYC pep4Δ top1Δ
TAP-SLD5 SEN1-9MYC pep4Δ tof1Δ
TAP-SLD5 SEN1-9MYC pep4Δ csm3Δ
50 min
20 min + SP
20 min
G1
Asyn

Figure 4.18. Sen1 interacts with replisome components in S phase independently of Ctf4. Cells carrying the *TAP-SLD5* and *SEN1-9MYC* alleles with or without deletion mutations of non-essential replisome components were grown in YPD to a density of 0.7 x 10^7 cells/ml in 1 l cultures at 30°C. The cells were arrested in G_1 using α-factor to a final concentration of 7.5 μg/ml for 3 h. After successful arrest, the cells were released from α-factor arrest by washing twice with fresh YPD, re-suspended in 1 l YPD and released in S phase at 30°C. The cells were harvested 20 min post-release and sample-prepared. FACS samples were taken for asynchronous cultures (Asyn), at G_1 arrest (G_1), at point of harvest in S phase (20 min), after preparation of S phase sample (20 min +SP) and 30 min after S phase harvest (50 min). The samples were used for IPs using TAP beads. **(A)** Cell extracts, IPs and FACS profiles of wildtype cells and cells carrying either the *ctf4Δ* , *tof1Δ* or *mrc1Δ* mutants. **(B)** Cell extracts, IPs and FACS profiles of wildtype cells and cells carrying either the *top1Δ*, *tof1Δ* or *csm3Δ* mutants. Strains used in this experiment: **(A)** CS**1125**, CS**1187**, CS**1217** and CS**1534 (B)** CS**1125**, CS**1217**, CS**1561** and CS**1676**.

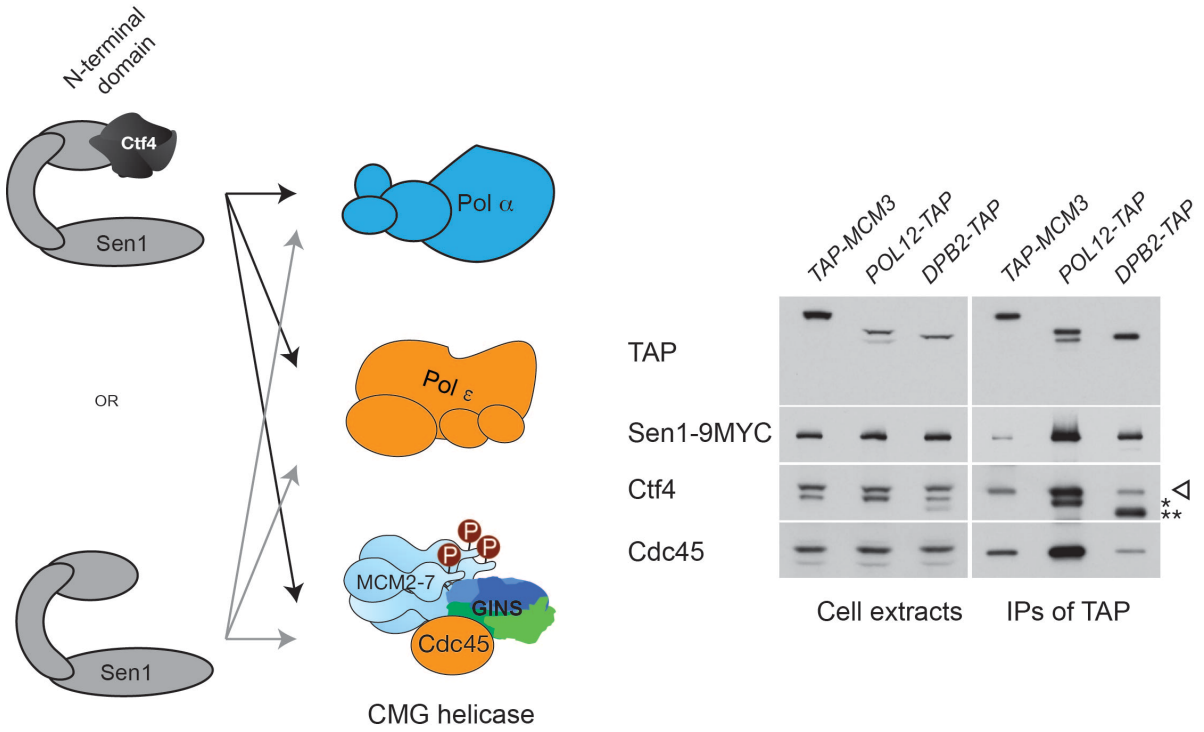


Figure 4.19. The ratio of Sen1 and Ctf4 levels co-precipitating with different sub-complexes of the replisome is not constant. Cells carrying the *SEN1-9MYC* allele and either one of *TAP-MCM3*, *POL12-TAP* or *DPB2-TAP* alleles were grown in YPD to a density of 0.7 x 10^7 cells/ml in 1 l cultures at 24°C. The cells were arrested in G_1 using α-factor to a final concentration of 7.5 μg/ml for 3 h. After successful synchronization, the cells were released from α-factor arrest by washing twice with fresh YPD, re-suspended in 1 l YPD and released in S phase at 24°C. The cells were harvested 30 min post-release and sample-prepared for use in IPs. Strains used in this experiment: CS**1134**, CS**1403** and CS**1416**. Note: The anti-Ctf4 antibody recognizes both Ctf4 specifically (arrow) and also recognizes the TAP tag non-specifically (marked as a single* or double stars**).

4.11 ABROGATION OF ORIGIN FIRING INTERFERES WITH THE INTERACTION OF SEN1 (2-931) WITH THE REPLISOME

CMG assembly requires activation of both the CDK and DDK kinases (Labib, 2010). Once activated, CDK phosphorylates Sld3 (and other proteins) and this enables recruitment of Cdc45 to licensed origins (Kamimura et al., 2001, Heller et al., 2011, Takayama et al., 2003, Tanaka et al., 2007). As such, degradation of Sld3 in G_1 prevents CMG assembly (and subsequent DNA replication) without interfering with other CDK-dependent cellular events.

Having established that Sen1 binds to at least another replisome component besides Ctf4, we wanted to determine whether the Sen1 (2-931) construct could co-purify with components of the replisome other than Ctf4 and GINS in the absence CMG assembly, but without interfering with the activity of either CDK and DDK. The Sen1 (2-931) construct was chosen instead of the full-length protein for two reasons. First, this construct is less labile than its full-length counterpart. This allows for harvesting of smaller volumes of culture (250 ml) and, hence, practical harvesting at different time-points within a single experiment. Secondly, given that the Sen1 (2-931) construct is both sufficient and necessary for binding to the replisome in S phase, it was deemed that the construct is an adequate substitute to the full-length protein as far as its interaction with replisome components is concerned.

In order to prevent CMG assembly, we used a temperature-degron of *SLD3*, *sld3-7-td* that can quickly be degraded upon shifting to non-permissive temperatures (37°C) (Kanemaki and Labib, 2006). We tested four strains: a control strain expressing *GAL1-TAP-Ø*, a positive control expressing *GAL1-TAP-SEN1 (2-931)* and two *sld3-7-td* strains expressing *GAL1-TAP-SEN1 (2-931)* and *GAL1-UBR1* either with or without *CTF4*. Cells were grown in YP medium supplemented with 2% (w/v) raffinose to a density of 0.7 x 10^7 cells/ml at a temperature of 24°C. In raffinose, neither the Sen1 construct nor the Ubr1 protein are expressed. The cultures were then synchronized in G_1 by addition of α-factor. Upon synchronization, half of the culture was harvested at −80°C.

The remainder of the cells was shifted to YPGAL supplemented with α-factor to maintain cells in G_1. The cells were left in YPGAL for 35 min to promote synthesis of both TAP-Sen1 (2-931) and Ubr1. Ubr1 is the E3 ligase required for degradation of proteins tagged with the TD degron. The cells were then shifted to 37°C for 1 h. This enables the degradation of Sld3-7 (Kanemaki and Labib, 2006). Finally, the cells were released in S phase (in YPGAL) at 37°C and were harvested 20 min post-release. Samples harvested both after synchronization in G_1 (in YP-Raff) and 20 min after release in S phase were used for TAP IPs. IPs of both the negative and positive controls correlated well with comparable experiments conducted at 24°C, suggesting that the interaction between the Sen1 construct and the replisome is not significantly affected at 37°C (Fig 4.20). However, in the absence of Sld3, Sen1 (2-931) was seen to co-precipitate only with GINS (this interaction with GINS is mediated by Ctf4) in S phase. As such, Sen1 (2-931) requires intact CMG complexes to bind to other component of the replisome. In the absence of both Sld3 and Ctf4, Sen1 (2-931) did not co-precipitate with any component of the replisome for which we tested.

These observations confirm that Sen1 can bind to GINS through Ctf4. However, from this experiment, we have also shown that when an intact CMG is present, Sen1 (2-931) can co-precipitate with replisome components besides Ctf4 and GINS. Absence of both Sld3 and Ctf4 would seem to completely abrogate Sen1's interaction with replisome components not only in S phase but throughout the cell cycle. In addition, this experiment suggests that the interaction between Sen1 and DNA polymerases is not simply dependent on the activity of CDK and DDK.

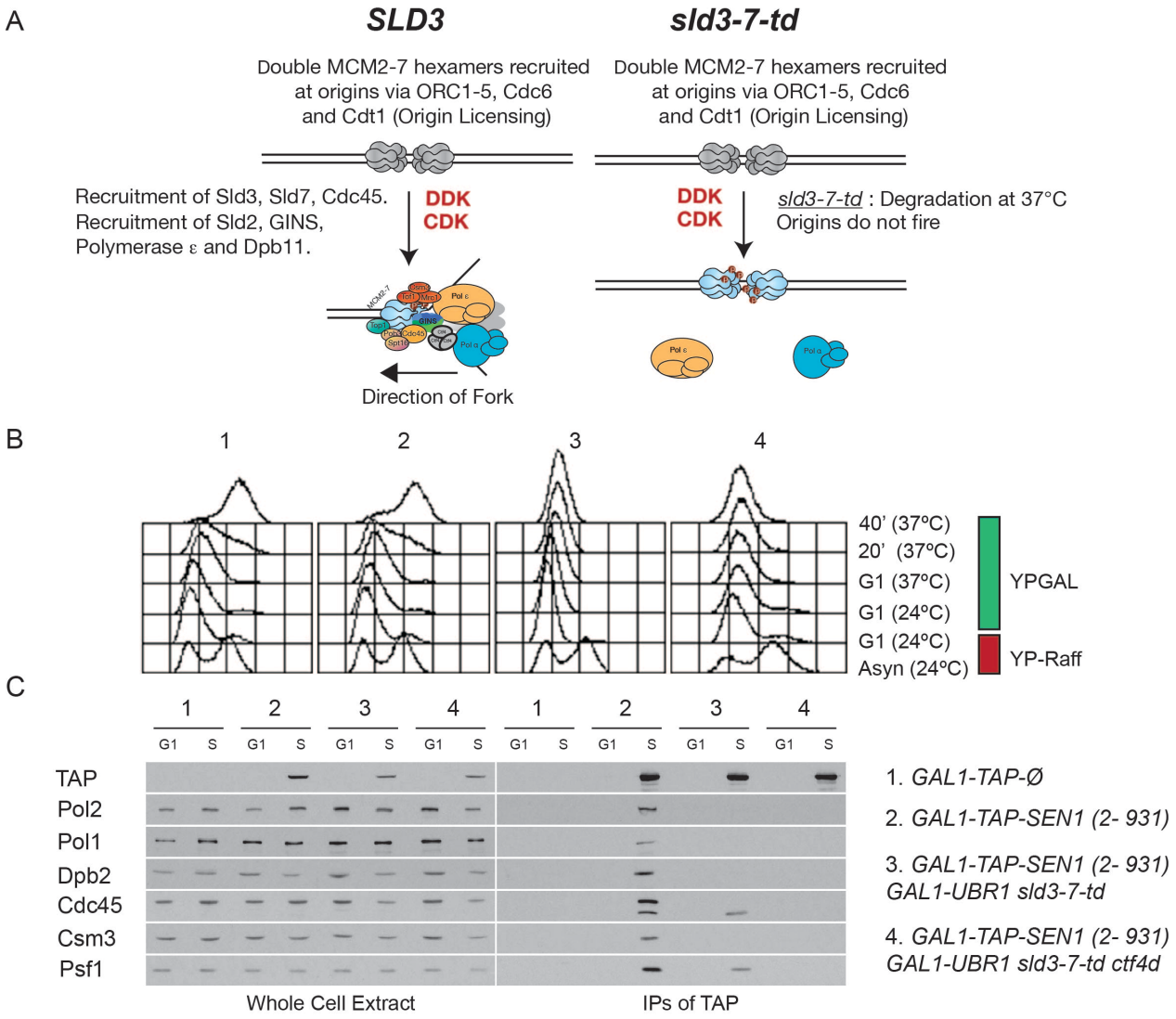

A
SLD3
sld3-7-td
Double MCM2-7 hexamers recruited at origins via ORC1-5, Cdc6 and Cdt1 (Origin Licensing)
Double MCM2-7 hexamers recruited at origins via ORC1-5, Cdc6 and Cdt1 (Origin Licensing)
Recruitment of Sld3, Sld7, Cdc45.
Recruitment of Sld2, GINS,
Polymerase ε and Dpb11.
DDK
CDK
DDK
CDK
sld3-7-td : Degradation at 37°C
Origins do not fire
Direction of Fork
B
1
2
3
4
40' (37°C)
20' (37°C)
G1 (37°C)
G1 (24°C)
G1 (24°C)
Asyn (24°C)
YPGAL
YP-Raff
C
1
2
3
4
1
2
3
4
G1
S
TAP
Pol2
Pol1
Dpb2
Cdc45
Csm3
Psf1
Whole Cell Extract
IPs of TAP
1. GAL1-TAP-Ø
2. GAL1-TAP-SEN1 (2- 931)
3. GAL1-TAP-SEN1 (2- 931)
GAL1-UBR1 sld3-7-td
4. GAL1-TAP-SEN1 (2- 931)
GAL1-UBR1 sld3-7-td ctf4d

Figure 4.20. Intact CMG is required for Sen1 (2-931) to interact with replisome components, besides Ctf4 and GINS. We grew cultures of four *MATa* strains: a negative control expressing *GAL1-TAP-Ø*, a positive control expressing *GAL1-TAP-SEN1 (2-931)* and two *sld3-7-td* strains expressing *GAL1-TAP-SEN1 (2-931)* and *GAL1-UBR1* with or without *CTF4*. Cells were grown in YP-Raff at 24°C to a density of 0.7 x 10^7 cells/ml. The cultures were then synchronized in G_1 by addition of α-factor to a final concentration of 7.5 µg/ml for 3 h. Half of the cells were then immediately harvested. The remainder of the cells was centrifuged out of solution, re-suspended in YPGAL supplemented with 7.5 µg/ml of α-factor and left to incubate at 24°C for 35 min to induce expression of both TAP-SEN1 (2-931) and Ubr1. Thereafter the cultures were transferred to 37°C for 1 h to degrade Sld3-7-TD to prevent CMG assembly. The cells were released in S phase by re-suspension in fresh YPGAL pre-warmed to 37°C. 20 min after release, the cells were harvested. The cultures were sampled for FACS analysis before addition of α-factor, upon synchronization in G1 (at first point of harvest), after induction of the TAP-SEN1 (2-931) and Ubr1, 1 h after shifting the cultures to 37°C and at 20 min (at second point of harvest) and 40 min after release in S phase. **(A)** Schematic of origin-licensing and firing in wildtype and *sld3-7-td* strains. **(B)** FACS profiles of the experiment. **(C)** Whole cell extracts and IPs samples. Strains used in this experiment: CS**1852** (1), CS**1957** (2), CS**2791** (3) and CS**2903** (4).

4.12 Residues 622-931 within the N-terminus of Sen1 is sufficient for its interaction with the replisome

Given that Sen1 seems to interact independently with at least two components of the replisome, we decided to focus on Sen1 itself to further characterize this interaction in order to eventually break it. Our results thus far suggest that mutants of Sen1 lacking the first 931 amino acids are unable to interact with replisome components. However, the *sen1-2* allele that lacks the first 945 amino acids displays a plethora of aberrant phenotypes, including mitochondrial defects (Sariki et al., 2016). Results here and elsewhere also indicate that first 931 residues of Sen1 are responsible for interaction with proteins not involved in DNA replication (Ursic et al., 2004, Nedea et al., 2008). Unsurprisingly then, ablation of the first 912 or 930 residues leads to slow-growth phenotypes, reminiscent of strains encoding the *sen1-2* allele (Fig 4.21). As such, simply removing the N-terminal domain of Sen1 is not a viable strategy to study the physiological importance of its interaction with the replisome.

For this reason, we wanted to identify the smallest domain of Sen1 within its N-terminal domain that is sufficient for its interaction with the replisome and we aimed to construct an allele of the gene lacking the corresponding bases. We cloned several constructs of Sen1 spanning residues 2-931 tagged at the N-terminus with 3HA (Fig S.2) and under control of the strong, inducible *GAL1* promoter at the *LEU2* locus in strains also carrying the *TAP-MCM3* allele. We then carried S phase TAP IPs for the different clones. *In silico* analysis of the N-terminal domain of Sen1 revealed that residues 913-931 house a coiled-coil domain. We wondered whether the latter was important for the interaction between Sen1 and the replisome. The Sen1 (410-913) construct was poorly expressed compared to the Sen1 (410-931) construct, suggesting that this coiled-coil domain was required to stabilize the protein. On the other hand, the results from the screen indicate that the first 621 amino acids of Sen1 are dispensable for its interaction with the replisome. Indeed, we found that a construct of Sen1 spanning residues 622-931 was sufficient to co-precipitate

with the replisome in S phase (Figs 4.22 and 4.23). However, attempts to use the Sen1 (761-931) construct in IPs was unsuccessful as the construct was poorly expressed, indicating that we had reached the technical limitations of the screen in trying to clone ever smaller fragments of Sen1 capable of co-precipitating with the replisome.

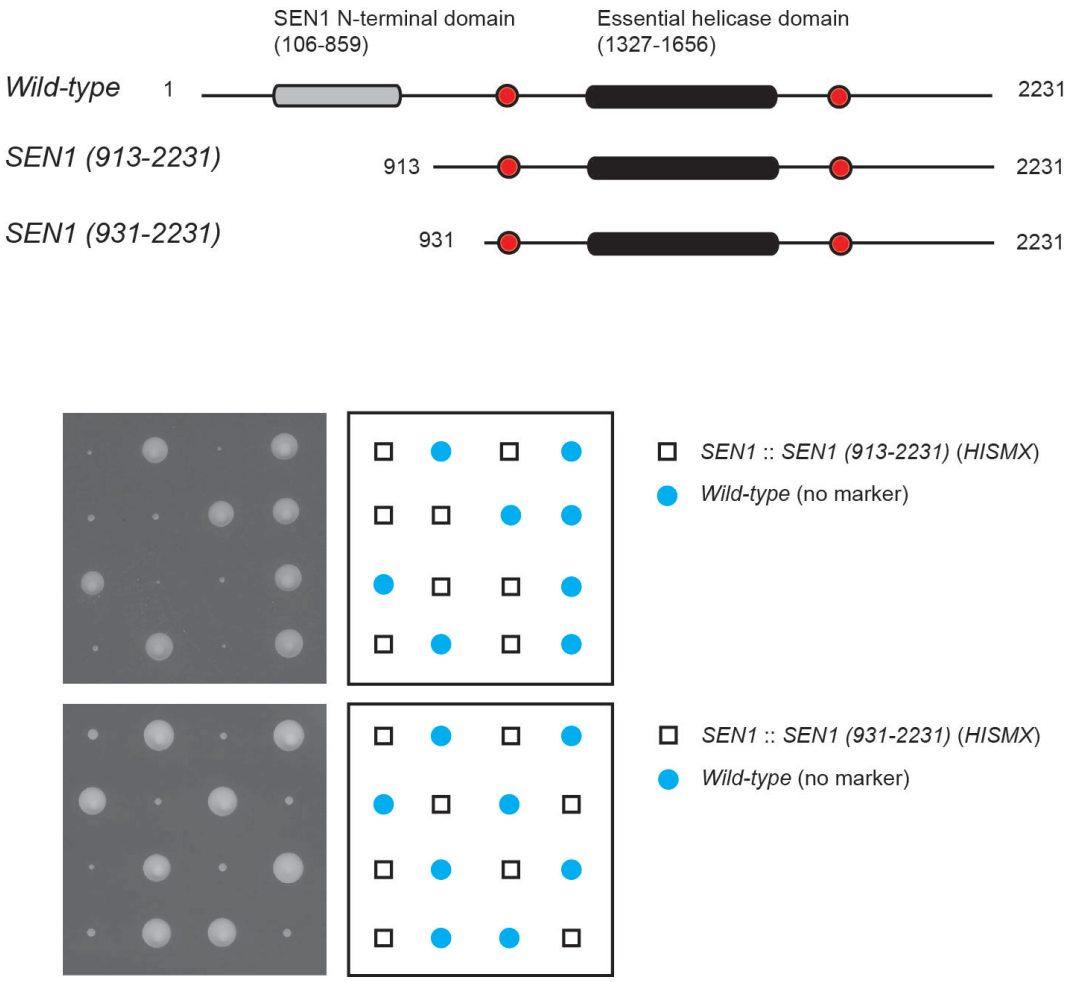


Figure 4.21. The *SEN1 (931-2231)* and *SEN1 (913-2231)* alleles lead to slow-growth phenotypes. Diploids of either *SEN1*/*SEN1 (931-2231)* or *SEN1*/ *SEN1 (913-2231)* were generated as described in Section 2.3.5. After sporulation, asci were treated with glucoronidase and individual asci were separated by microscopic manipulation into their constituent spores onto solid YPD medium. The spores were allowed to germinate at 30°C and scored by growth on selective media. Strains used in the experinent: CS**2457** and CS**2458**. Donor strains: CS**2403** and CS**2404**.

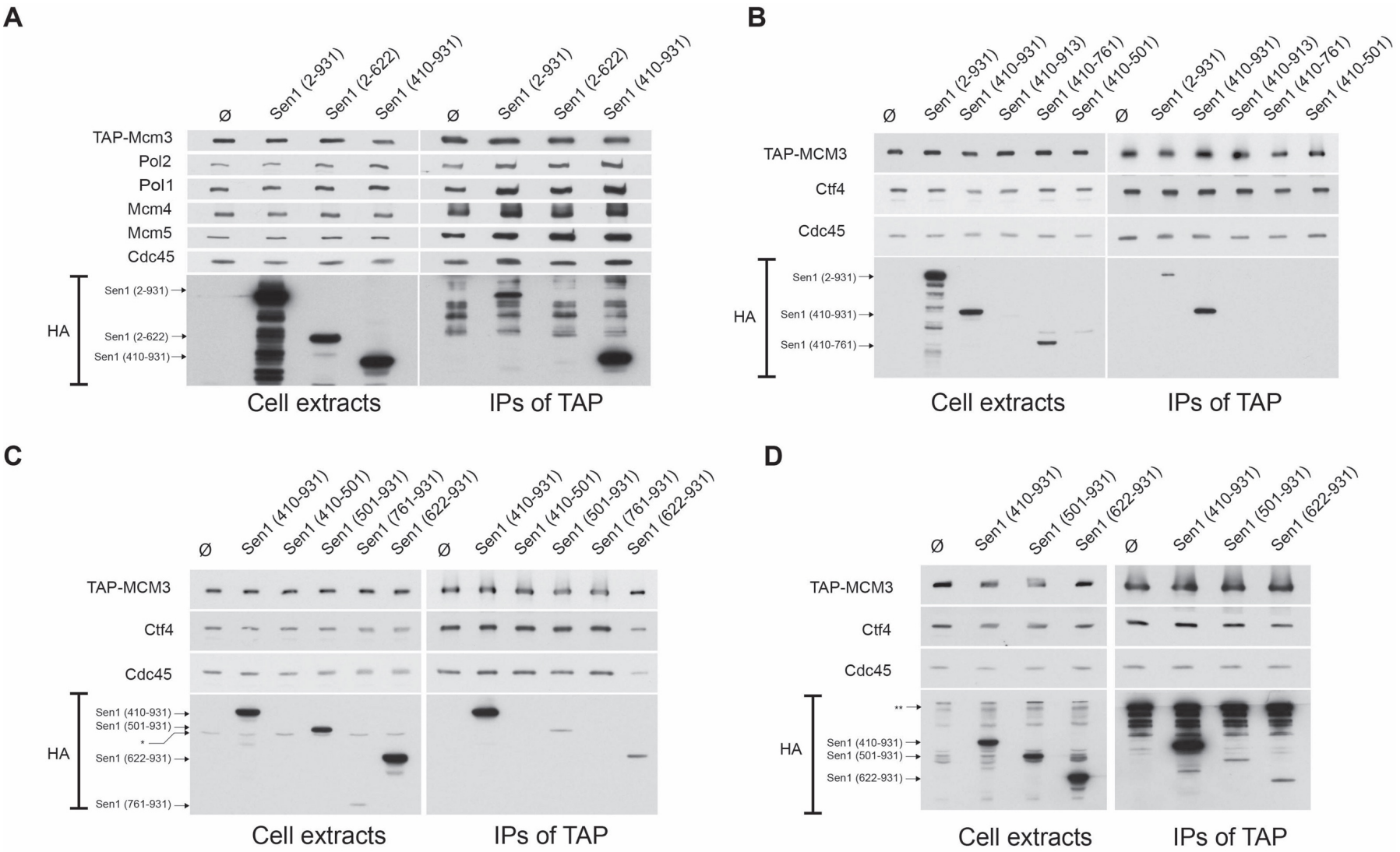

A
Ø
Sen1 (2-931)
Sen1 (2-622)
Sen1 (410-931)
TAP-Mcm3
Pol2
Pol1
Mcm4
Mcm5
Cdc45
HA
Cell extracts
IPs of TAP
B
Sen1 (410-913)
Sen1 (410-761)
Sen1 (410-501)
TAP-MCM3
Ctf4
Cdc45
Sen1 (410-761)
C
Sen1 (410-931)
Sen1 (410-501)
Sen1 (501-931)
Sen1 (761-931)
Sen1 (622-931)
*
Sen1 (761-931)
D
**

Figure 4.22. Residues spanning 622-931 within the N-terminus of Sen1 are sufficient for interaction with the replisome. Cells carrying the *TAP-MCM3* allele and different N-terminally tagged Sen1 fragments under the *GAL1* promoter were grown in YP-Raff at 24°C to a density of 0.7 x 10^7 cells/ml in 250 ml cultures. The cells were arrested in G_1 using α-factor to a final concentration of 7.5 μg/ml for 3 h. After successful arrest, YPGAL medium was substituted for YP-Raff, keeping α-factor at a concentration of 7.5 μg/ml for a further 35 min. The cells were then released from G_1 arrest by washing out the α-factor with fresh YPGAL, followed by release in S phase in YPGAL at 24°C. Cells were harvested 30 min after release. The harvests were used for IPs using TAP beads. **(A- D)** Westerns of cell extracts and IPs of TAP for four different experiments. Strains used in this experiment: CS**1711**, CS**1714**, CS**2030**, CS**2032**, CS**2145**, CS**2146**, CS**2147**, CS**2148**, CS**2149** and CS**2150**. Two non-specific bands are observable upon blotting against the HA tag. These are indicated as either * or **.

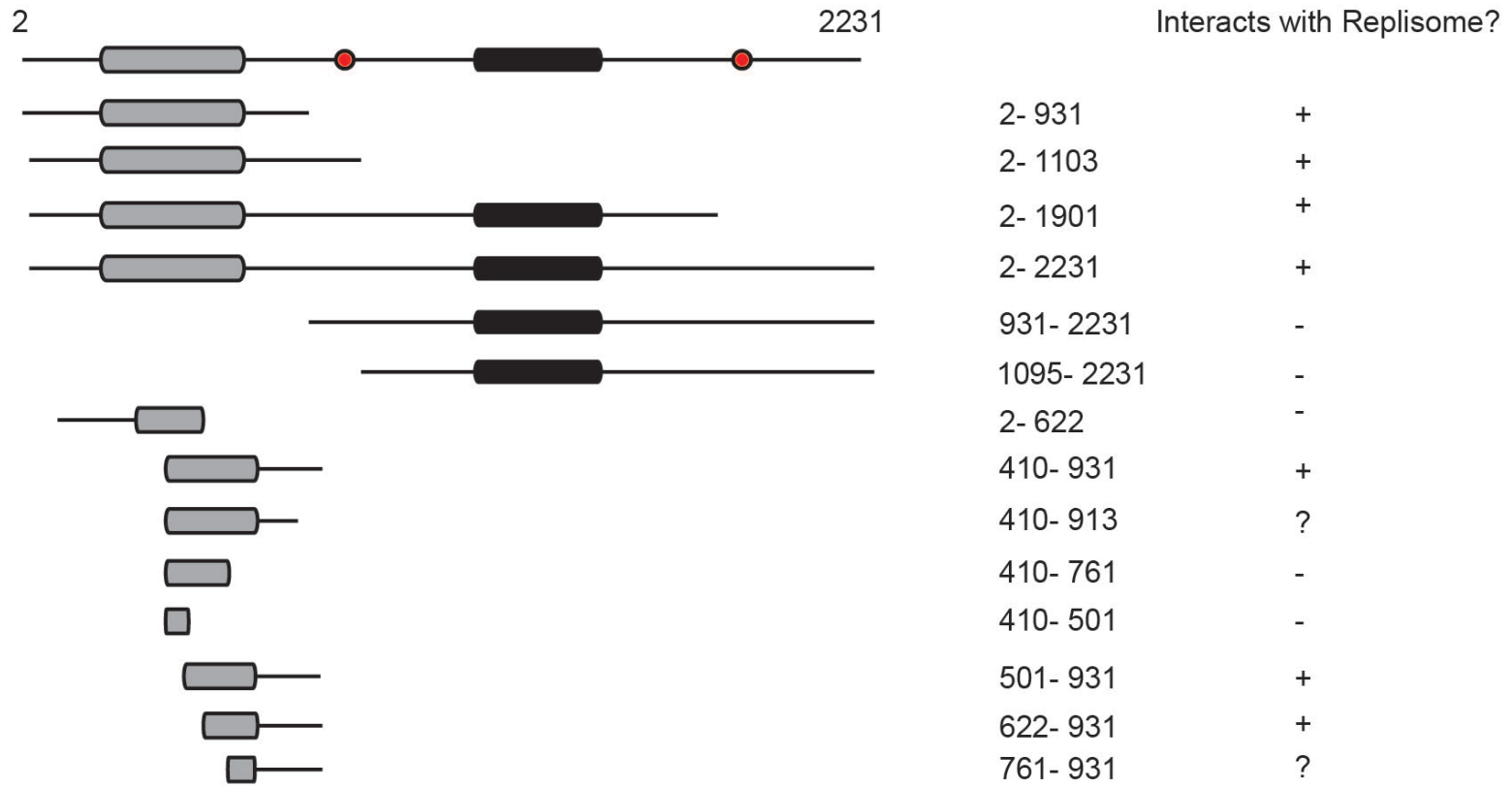


Figure 4.23. Schematic summarizing the different constructs of Sen1 tested and their abilities to interact with replisome components in S phase.

4.13 MINI-TRUNCATIONS WITHIN THE N-TERMINUS OF SEN1 LEAD TO SYNTHETIC DEFECTS

Having established that residues 622-931 were sufficient for Sen1's interaction with the replisome, we then designed the *SEN1 (410-931Δ)-TAP* and *SEN1 (622-931Δ)-TAP* alleles in an attempt to construct alleles of *SEN1* that encode for proteins that do not bind to the replisome whilst leaving large portions of the N-terminal domain untouched.

The cloning was done in a diploid with one of the copy of *SEN1* previously disrupted with the *URA3* gene (Section 2.3.5). Integration of PCRs simultaneously substituted the *URA3* gene with the novel allele of *SEN1* whilst also tagging it with the TAP tag at its C-terminus and marking it with the reporter *kanMX* gene. The resulting clones were grown on FOA plates to select for *URA* auxotrophy. Colonies were then selected for growth on YPD supplemented with kanamycin and screened for integration at the correct locus by PCR. Individual clones were also screened by Western blotting. In brief, had the PCR integrated correctly, the TAP signal would be visible for the clone showing that there were no frameshifts introduced. Additionally, the TAP signal would run below that of a control Sen1-TAP clone. Clones that were judged to have integrated correctly were sequenced for the *SEN1* gene and if the correct truncation mutation was present, the clones would be transferred onto sporulation medium and the resulting tetrads scored.

Surprisingly, the *SEN1 (410-931Δ)-TAP* and *SEN1 (622-931Δ)-TAP* alleles caused a greater defect compared to either *SEN1 (913- 2231)* or *SEN1 (931-2231)* (Fig 4.24). In fact, the *SEN1 (410-931Δ)* allele led to micro-colonies at 30°C whilst cells encoding the *SEN1 (622-931Δ)* allele were distinguished by terminal arrest in G_2 after germination, similar to *sen1Δ* cells. This suggests that the additional amino acid regions render *SEN1* mutants unstable compared to *SEN1 (931-2231)* making them unsuitable mutants. We also tried to bridge the 409th or 623rd residues to the 932nd residue with either a small or

long linker or with a region of DNA coding for the Ctf4 Interacting Peptide from Pol1 (CIP). None of these strategies led to viable mutants (not shown).

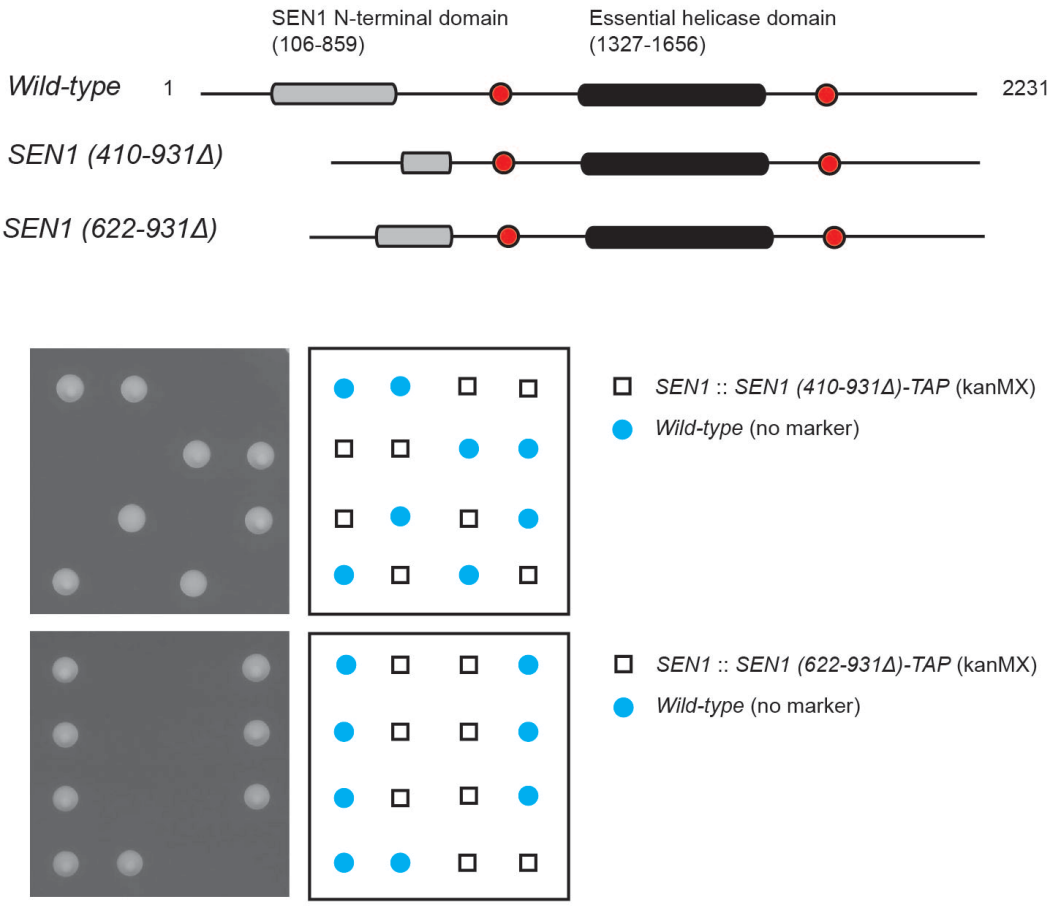


Figure 4.24. The *SEN1 (410-931Δ)-TAP* and SEN1 *(622-931Δ)-TAP* alleles lead to micro-colonies and terminal arrest in G_2 respectively. Diploids of either *SEN1*/*SEN1 (410-931Δ)-TAP* or *SEN1*/*SEN1 (622-931Δ)-TAP* were generated as described in Section 2.3.5. After sporulation, asci were treated with glucoronidase and individual asci were separated by microscopic manipulation into their constituent spores onto solid YPD medium. The spores were allowed to germinate at 30°C and scored by growth on selective media. Donor strains: CS**2334** and CS**2335**.

4.14 IDENTIFYING POINT MUTANTS TO DISRUPT THE INTERACTION BETWEEN SEN1 AND THE REPLISOME

The residues spanning 622-931 in Sen1 did not reveal any CIP box that could help mediate the interaction between Sen1 and Ctf4 (Samora et al., 2016). However, secondary structure analysis of those residues using the Jpred4 software (Drozdetskiy et al., 2015) predicted several α-helices and, accordingly, several residues were predicted to be buried with at most 25% solvent accessibility. The Jpred4 software uses multiple-alignment of the input sequence (here residues spanning 501-1200 in Sen1) with several of its close orthologues to predict secondary structures. Moreover, analysis of the protein structure (using residues spanning 622-931) by the Phrye2 suggested that the N-terminal domain of Sen1 contained several Armadillo (ARM) repeats, similar to the human β-catenin protein (Kelley et al., 2015). ARM repeats are characterized by degenerate primary sequences but conserved tertiary structures made of three α-helices spanning around 40 residues collectively. Several ARMs fold and interact with one another in such a way as to produce a rigid super-helix of α-helices (Fig 4.25). The inner concave surface of the super-helix is positively-charged and it can promote protein-protein interactions (Tewari et al., 2010). To construct a mutant of Sen1 that is incapable of interacting with the replisome, we chose to conduct a screen of Sen1 mutants where residues of the protein that are both conserved (Fig 4.26) and that are predicted to occur on the surface of the super-helix (Yachdav et al., 2014) were mutated. The Jpred4 software was used to identity the various residues predicted to occur on the surface. These have a predicted solvent accessibility above 25%. We did so by adopting a site-directed mutagenesis approach (Qiagen). The residues targeted are depicted in Table 4.3.

Table 4.3. The residues that were targeted for mutagenesis to disrupt Sen1's interaction with the replisome is depicted here.

Mutant name	Point Mutations
sen1-3	W773A E774A W777A
sen1-4	L656A S657A K658A I659A L660A
sen1-5	D850A E851G V852A L853G L854A
sen1-6	V746G D747G P748G I749G

For the screen, we first cloned the *3HA-SEN1 (2-2231)* allele at an ectopic locus (*LEU2*) under its own promoter whilst the endogenous *SEN1* allele was deleted. We found that, at the ectopic locus, the *3HA-SEN1 (2-2231)* was able to counter the lethal phenotype associated with *sen1*Δ (Fig 4.27). However, these cells grew poorly, suggesting the ectopic allele cannot fully complement *sen1*Δ*.* Cloning the *3HA-SEN1 (931-2231)* allele at the ectopic locus led to spores that germinate but terminally arrest in G_2, similar to *sen1*Δ*.* Thus, this allele cannot complement *sen1*Δ when cloned under the *SEN1* promoter at the *LEU2* locus. However, we have previously cloned cells with the *SEN1 (931-2231)* allele at its endogenous locus and these cells were sick but viable. Taken together, these observations suggest that either the 3HA tag or an incomplete *SEN1* promoter was introducing a synthetic defect. We then substituted the strong constitutive *ACT1* promoter for the *SEN1* promoter. Under the *ACT1* promoter, *3HA-SEN1 (2-2231)* cloned at the *LEU2* locus was able to fully complement *sen1*Δ and cells carrying the *3HA-SEN1 (931-2231)* allele at the ectopic locus in *sen1*Δ cells behaved similarly to cells carrying the *SEN1 (931-2231)* allele cloned at its endogenous locus. Thus, we chose to use *SEN1* alleles cloned under the *ACT1* promoter for our screen. Upon cloning of the mutant *SEN1* alleles, we found that none of the mutants generated displayed any significant growth defects (Fig 4.28).

We then carried out S phase IPs in strains carrying the *TAP-MCM3* allele along with wildtype *3HA-SEN1 (2-2231)* or the novel mutant alleles. A strain carrying an untagged *SEN1* allele was used as a negative control. We found that cells carrying the *sen1-3* allele did not co-precipitate with the replisome

(Fig 4.29). The *sen1-3* allele is characterized by the W773A E774A W777A substitutions and the two mutated tryptophans correspond to the most conserved residues within the ARM repeats (Fig 5.7). Also notable was that MCM2-7 IPs in strains carrying the *sen1-4*, *sen1-5* and *sen1-6* alleles were characterized by stronger interaction between the replisome and Sen1. This could arise by morphing of the surface of the super-helix in such a way as to alter the strength of protein-protein interactions. As such, the strategy utilized in this study could be used as a blueprint to study interactions between ARM repeats proteins and their binding partners. The levels of the Sen1-3 protein were lower than that of its wildtype counterpart. This was also true for the Sen1-4 and Sen1-5 proteins. This could indicate that these mutants are less stable than wildtype Sen1 (the mutations are unlikely to affect transcription downstream of the *ACT1* promoter). It is possible that the reduced levels of Sen1-3 could contribute to its lowered affinity for replisome components. On the other hand, Sen1-4 and Sen1-5 seem to interact more strongly with replisome components. As such any decreased stability of the Sen1-3 mutant cannot, on its own, account for the lowered affinity of the mutant for replisome components.

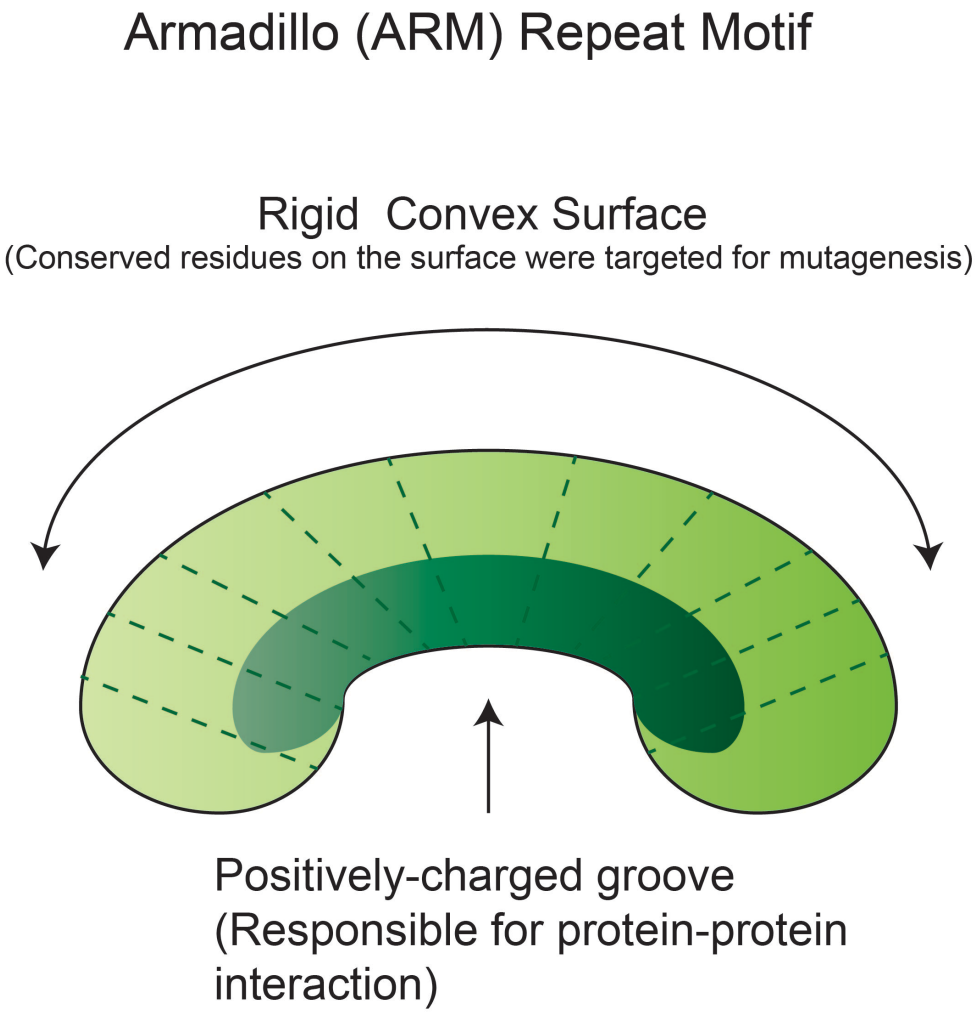


Figure 4.25. Armadillo repeats fold and interact with one another to form a super-helix of α-helices. The positively charged groove at its concave surface in thought to be responsible for protein-protein interactions. Importantly, the N-terminal domain of Sen1 does not contain a CIP box, but instead is predicted to encode for armadillo repeats.

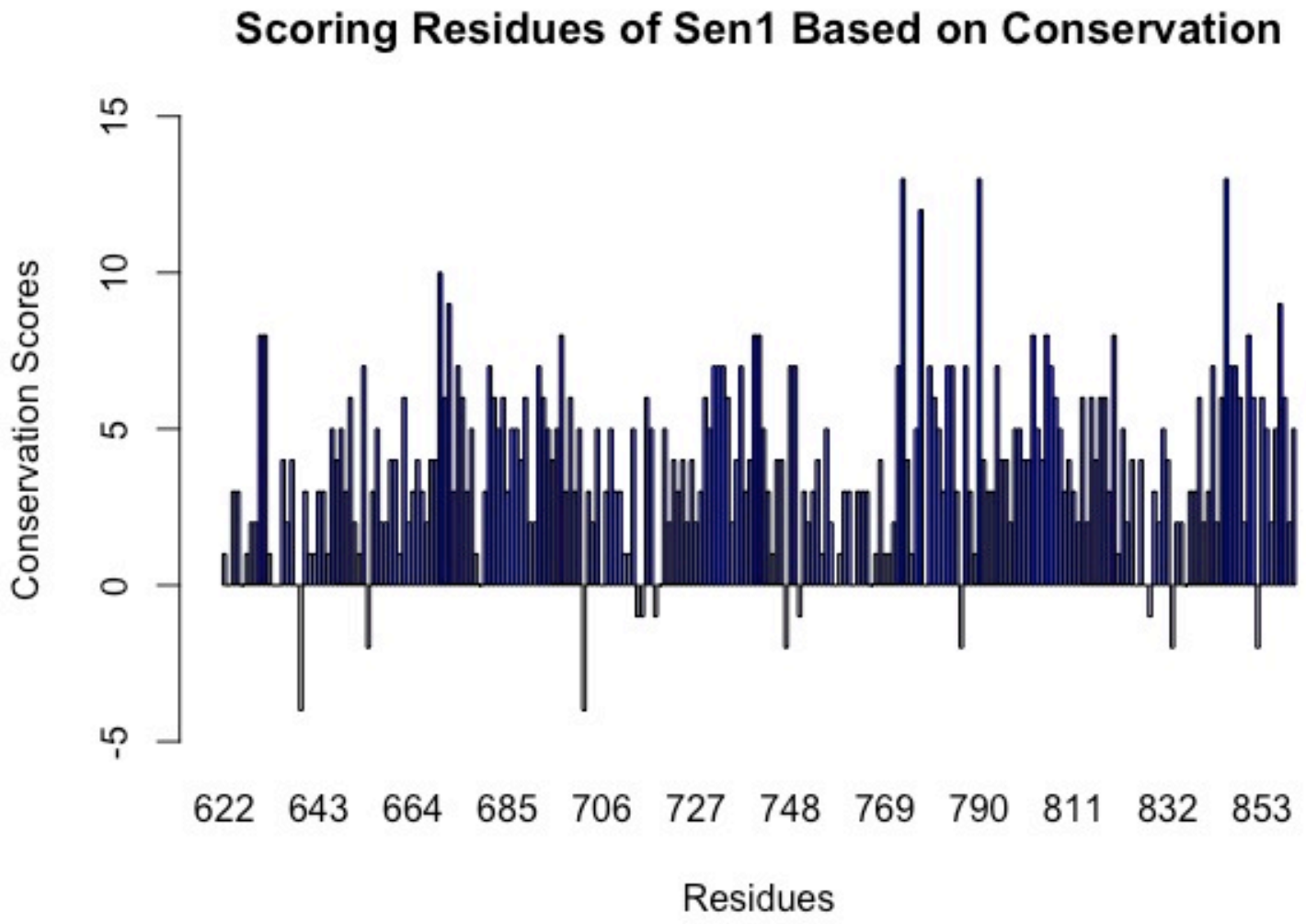

Scoring Residues of Sen1 Based on Conservation
Conservation Scores
15
10
5
0
-5
622
643
664
685
706
727
748
769
790
811
832
853
Residues

Figure 4.26. Conservation scores of Sen1 residues spanning 622-931 against a consensus sequence. Armadillo repeats are conserved structurally but show less conservation at the primary sequence. Nonetheless, by comparing close orthologues of Sen1 (via BLASTP), a consensus sequence was built against which the different residues of Sen1 (622-931) was be scored for conservation. In brief, residues 622-931 was analysed by PSI-BLAST (Position-Specific Iterative Basic Local Alignment Search Tool) (Altschul et al., 1997). The first iteration of the algorithm works as a normal BLASTp where the query amino acid sequence is compared to protein sequences in an unbiased database in order to identify regions of local alignment. The output of this first iteration is a multiple alignment of the highest scoring pairs as well as a position specific score matrix (PSSM) or profile generated from the multiple alignment of the query sequence. The PSSM captures the extent of conservation at individual amino acid positions. The second iteration of the algorithm then uses this PSSM as a query sequence against an unbiased database. Sequences that score above a set threshold are added to the multiple alignment and the PSSM is refined accordingly. Subsequent iterations are then used until no new sequences are added to the multiple alignment. This leads to generation of a final PSSM to which the residues of the original query sequence (here, Sen1 (622-931)) are scored. The higher the score, the more conserved are the individual residues.

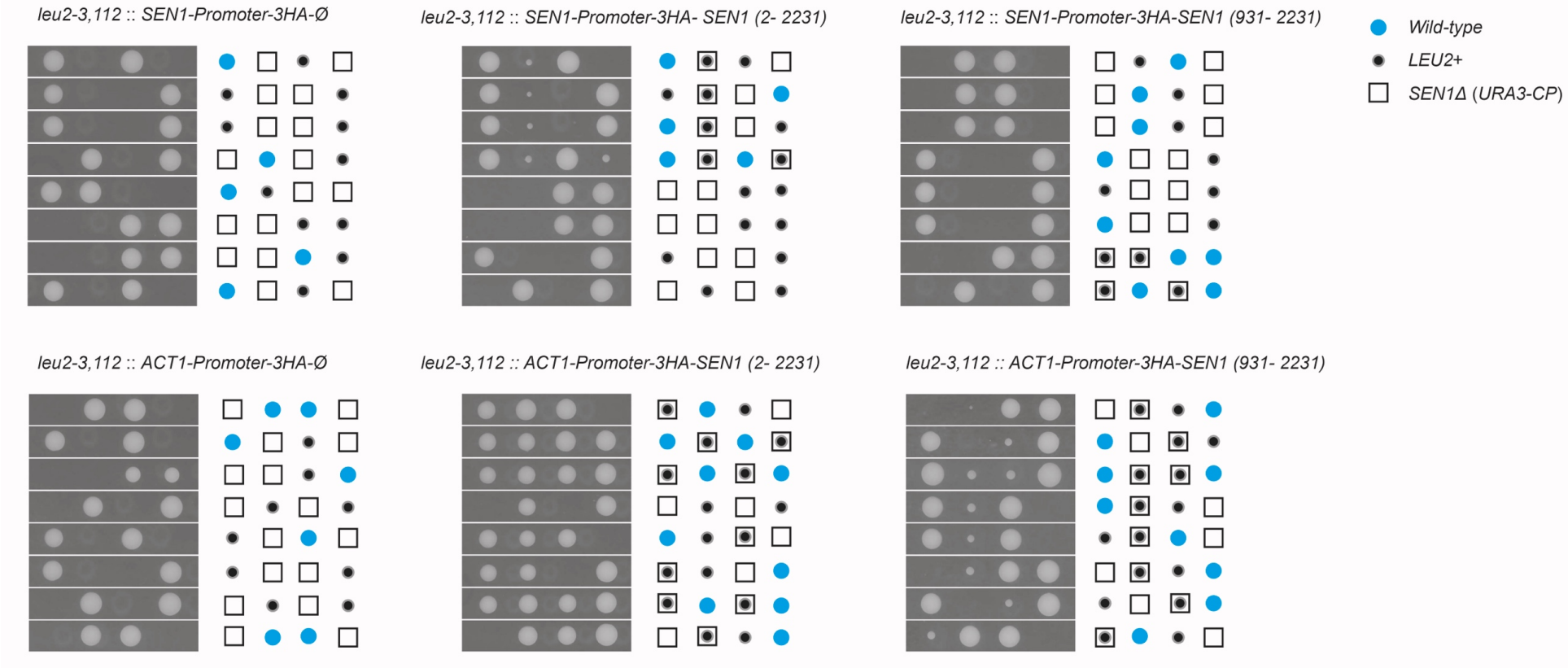


Figure 4.27. Expressing *SEN1* at an ectopic locus. Viable strains derived from the dissections: CS**2582**, CS**2584**, CS**2586**.

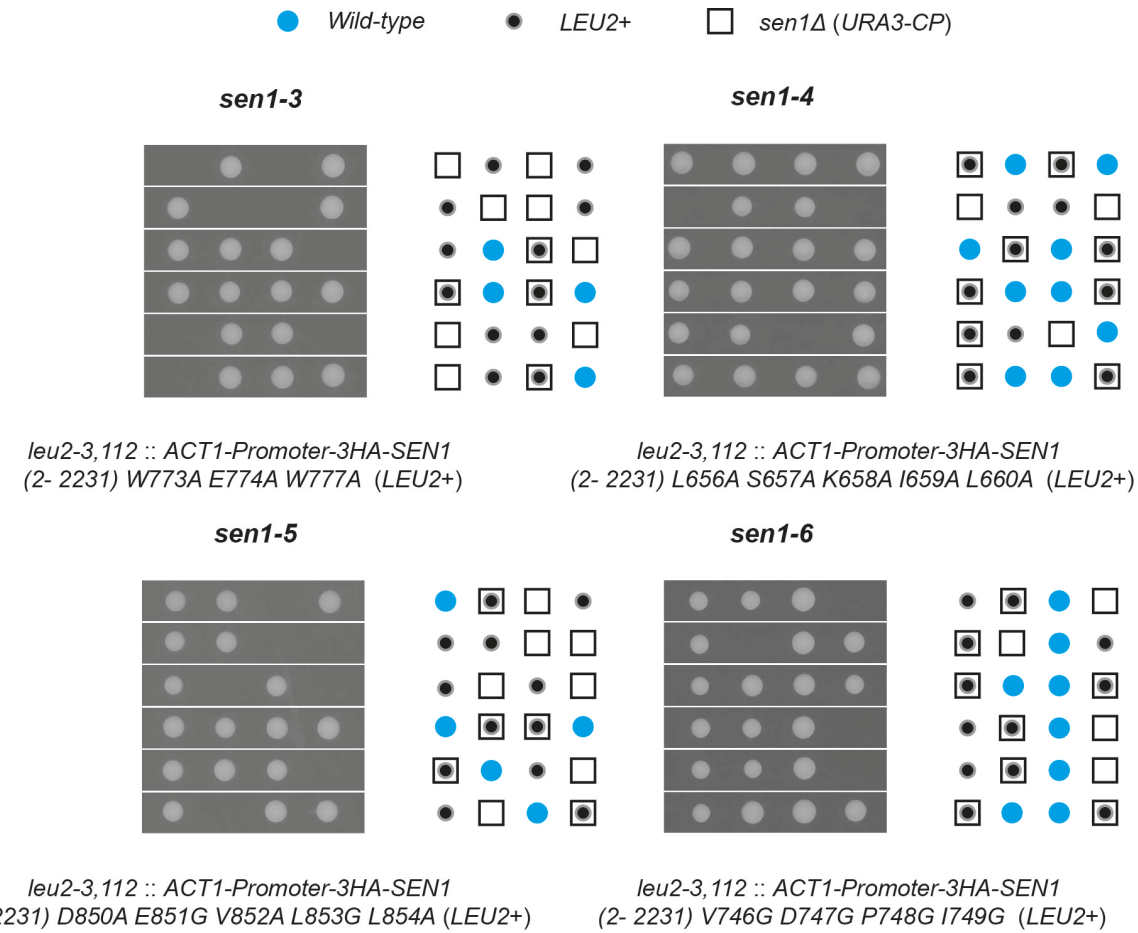


Figure 4.28. Mutants of *SEN1* used in the screen. None of the four mutants used to screen for loss of interaction with the replisome presented any significant growth defects and were able to complement *sen1*Δ fully. Viable strains derived from the dissections: CS**2607**, CS**2609**, CS**2617**, CS**2623**.

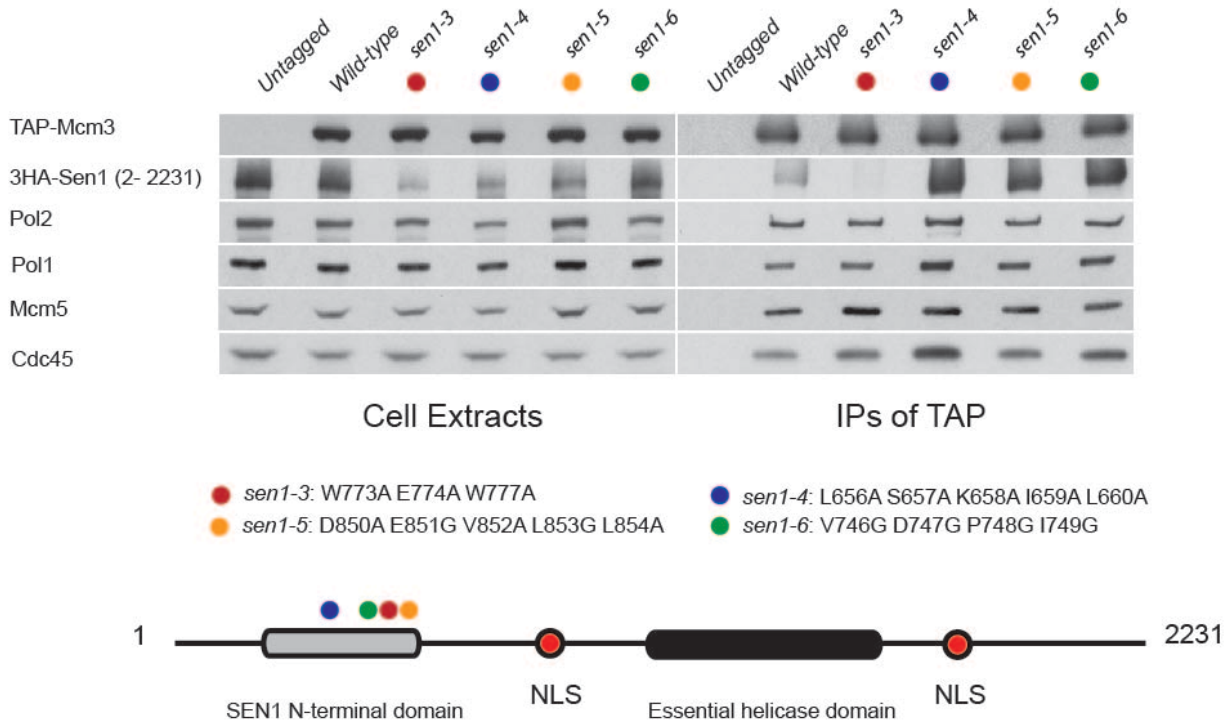


Figure 4.29. Novel alleles of *SEN1* abrogate or reinforce the interaction of the protein with the replisome. Cells carrying the *TAP-MCM3* allele with novel alleles of *SEN1* tagged at the N-terminus with the 3HA tag were grown in YPD to a density of 0.7 x 10^7 cells/ml in 1 l cultures at 24°C. The cells were arrested in G_1 using α-factor to a final concentration of 7.5 μg/ml for 3 h. After successful arrest, the cells were released from α-factor arrest by washing twice with fresh YPD, re-suspended in 1 l YPD and released in S phase at 24°C. The cells were harvested 30 min post-release and sample-prepared for use in IPs. Strains used in this experiment: CS**2636**, CS**2638**, CS**2640**, CS**2642**, CS**2669** and CS**2670**. In this particular experiment, the protein levels of Sen1-3, Sen1-4 and Sen1-5 appear lower than that of the wildtype protein, as observed on Western blots.

4.15 BIOLOGICAL CONSEQUENCES OF ABROGATING THE INTERACTION BETWEEN SEN1 AND THE REPLISOME

Having established biochemically that Sen1-3 is defective for interaction with the replisome, we then set out to characterize the phenotypic consequences of such a loss of interaction *in vivo*. In isolation, none of the novels alleles of *SEN1* were especially sensitive to non-optimal temperatures (25°C or 37°C), compared to wildtype. This contrasts to cells carrying the *ACT1-3HA-SEN1 (931-2231)* allele that are slow-growing at both 25°C and 37°C (Fig 4.30). Nor did we detect any slow growth on non-fermentable sources of carbon (either glycerol or ethanol) that is characteristic of mitochondrial defects. Again, this contrasts with cells carrying the *ACT1-3HA-SEN1 (931-2231)* allele. Growth in either the dNTP-depleting agent hydroxyurea (HU) or the alkylating agent methylmethanosulfate (MMS) did not reveal any defect of the novel alleles of Sen1 compared to wildtype cells (Fig 4.30). Taken together, these results suggest that the point mutations used here do not grossly affect the functionality of the N-terminal domain of the protein.

Given that depletion of Sen1 leads to increased levels of R-loops (Mischo et al., 2011), we wondered whether the *sen1-3* mutant interacted genetically with mutants of proteins that either prevent the formation of R-loop (such as Top1) or with mutants that actively remove R-loops (such as the RNase H enzymes) (Aguilera and García-Muse, 2012). As a control, the other novel alleles of *SEN1* were also used to detect synthetic interactions. Cells carrying the novel *SEN1* alleles were not characterized by noticeable growth defects in the absence of Top1 (Fig 4.30). On the other hand, the *sen1-3* allele conferred temperature-sensitivity at 37°C when grown in cells deleted for both *RNH1* and *RNH201*, demonstrating a synthetic defect that is not observable in *rnh1Δ rnh201Δ* cells. Moreover, the *sen1-3* allele conferred increased susceptibility to both HU and MMS in *rnh1Δ rnh201Δ* cells (Fig 4.30). Given the role of RNases H1 and H2 in the removal of R-loops, these observations suggest that the association of Sen1 with replisomes helps cells to efficiently remove R-loops. It is known that failure of nascent mRNA to be processed via the

THO/TREX complex also leads to stable R-loops (Huertas and Aguilera, 2003). To confirm whether the synthetic defects in *rnh1Δ rnh201Δ sen1-3* cells were indicative of stabilized R-loops, we crossed the *sen1-3* allele into *hrp1Δ* cells. Hrp1 is a component of the THO/TREX complex. This double mutant also showed increased susceptibility to both HU and MMS, again suggesting that cells carrying the *sen1-3* allele are not fully proficient in R-loop removal. Taken together, this suggests that loss of interaction between Sen1 and the replisome in *sen1-3* cells leads to more stable R-loops.

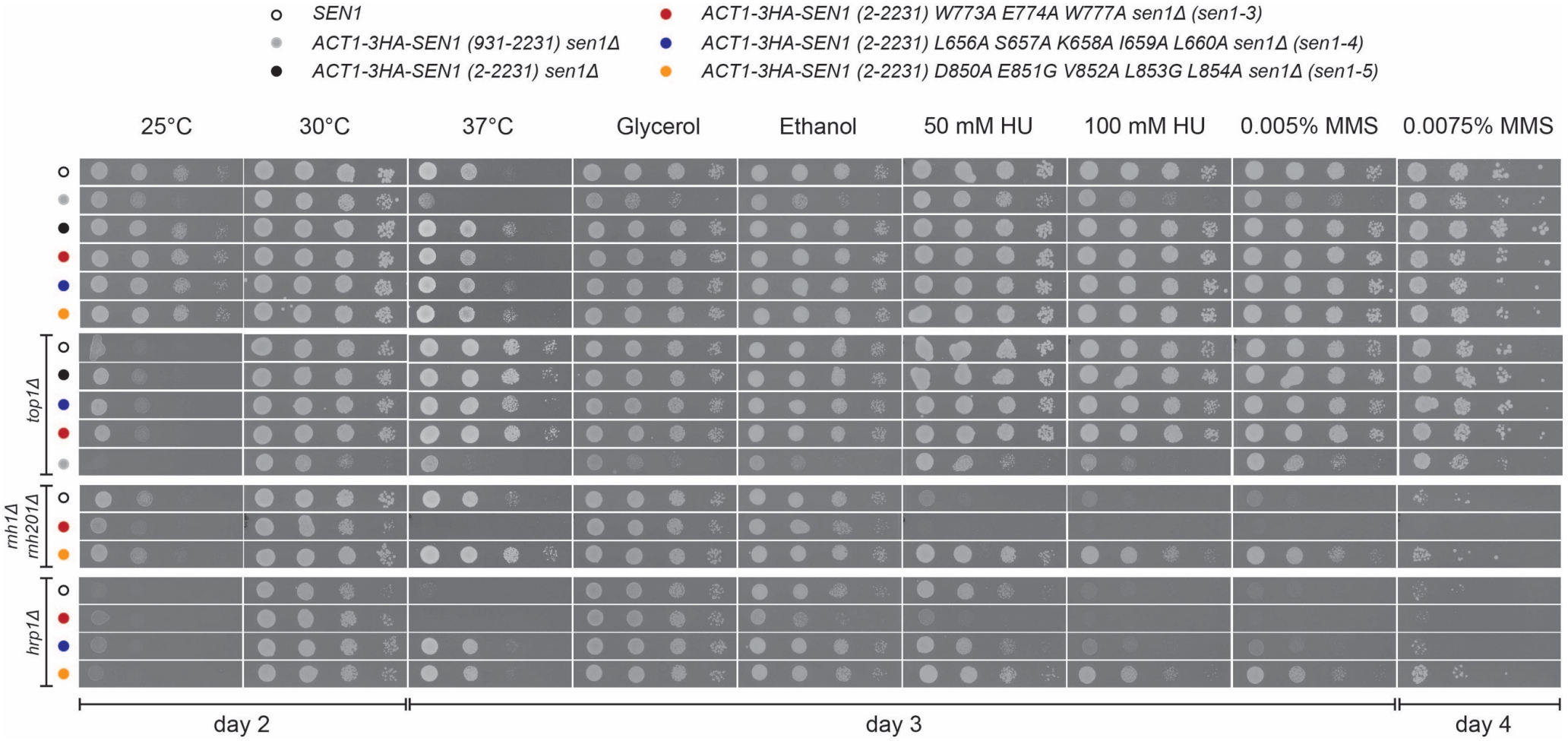

SEN1
ACT1-3HA-SEN1 (931-2231) sen1Δ
ACT1-3HA-SEN1 (2-2231) sen1Δ
ACT1-3HA-SEN1 (2-2231) W773A E774A W777A sen1Δ (sen1-3)
ACT1-3HA-SEN1 (2-2231) L656A S657A K658A I659A L660A sen1Δ (sen1-4)
ACT1-3HA-SEN1 (2-2231) D850A E851G V852A L853G L854A sen1Δ (sen1-5)
25°C
30°C
37°C
Glycerol
Ethanol
50 mM HU
100 mM HU
0.005% MMS
0.0075% MMS
top1Δ
rnh1Δ rnh201Δ
hrp1Δ
day 2
day 3
day 4

Figure 4.30. The novel alleles of *SEN1* show synthetic defects in the absence of *RNH1* and *RNH201*. Strains were grown from frozen on fresh YPD at 30°C for 2 days. The cells were then diluted in water to obtain cell suspensions of 5 x10^6, 5 x 10^5, 5 x10^4 and 5 x10^3 cells/ml. 10 μl of each suspension was pipetted onto either non-selective medium or non-selective medium supplemented with either HU or MMS. Cells were grown up to 4 days and imaged daily. For each condition, images from the same day are presented. Strains used (top to bottom): (topmost panel) CS**1**, CS**2582**, CS**2584**, CS**2607**, CS**2623**, CS**2609**; (second panel from top) CS**44**, CS**2656**, CS**2659**, CS**2661**, CS**2668**; (third panel from top) CS**2734**, CS**2736**, CS**2738**; (bottom panel) CS**2276**, CS**2696**, CS**2702** and CS**2729**.

Given that we had cloned the novel alleles of *SEN1* under the strong and constitutive *ACT1* promoter, the increased levels of the Sen1 molecules could mask some of the defects associated with the *sen1-3* allele. To test this possibility, we generated the *sen1-3* mutation at the endogenous *SEN1* locus. We also tagged the gene with the TAP-tag. Using this strain, we carried out S phase IPs of TAP and found that Sen1-3-TAP, unlike its wildtype counterpart, did not co-purify with replisome components in S phase, including Ctf4 (Fig 4.31). As such, cloning *sen1-3* at its endogenous locus recaptures the loss of interaction when the allele is cloned at an ectopic site. We then constructed triple mutants of *rnh1Δ rnh201Δ sen1-3-TAP*. As a control, we found that cells carrying the *rnh1Δ rnh201Δ SEN1-TAP* alleles grew at similar rates to cells carrying only *rnh1Δ rnh201Δ* alleles, suggesting that the C-terminal TAP tag did not affect growth. By contrast, the *rnh1Δ rnh201Δ sen1-3-TAP* genotype was inviable (Fig 4.32). This confirms the genetic interaction between the RNase H enzymes and Sen1, suggesting that Sen1's interaction with the replisome is required for the removal of DNA/RNA hybrids. In addition, overexpressing *sen1-3* under control of the *ACT1* promoter can mask some of defects linked to this allele.

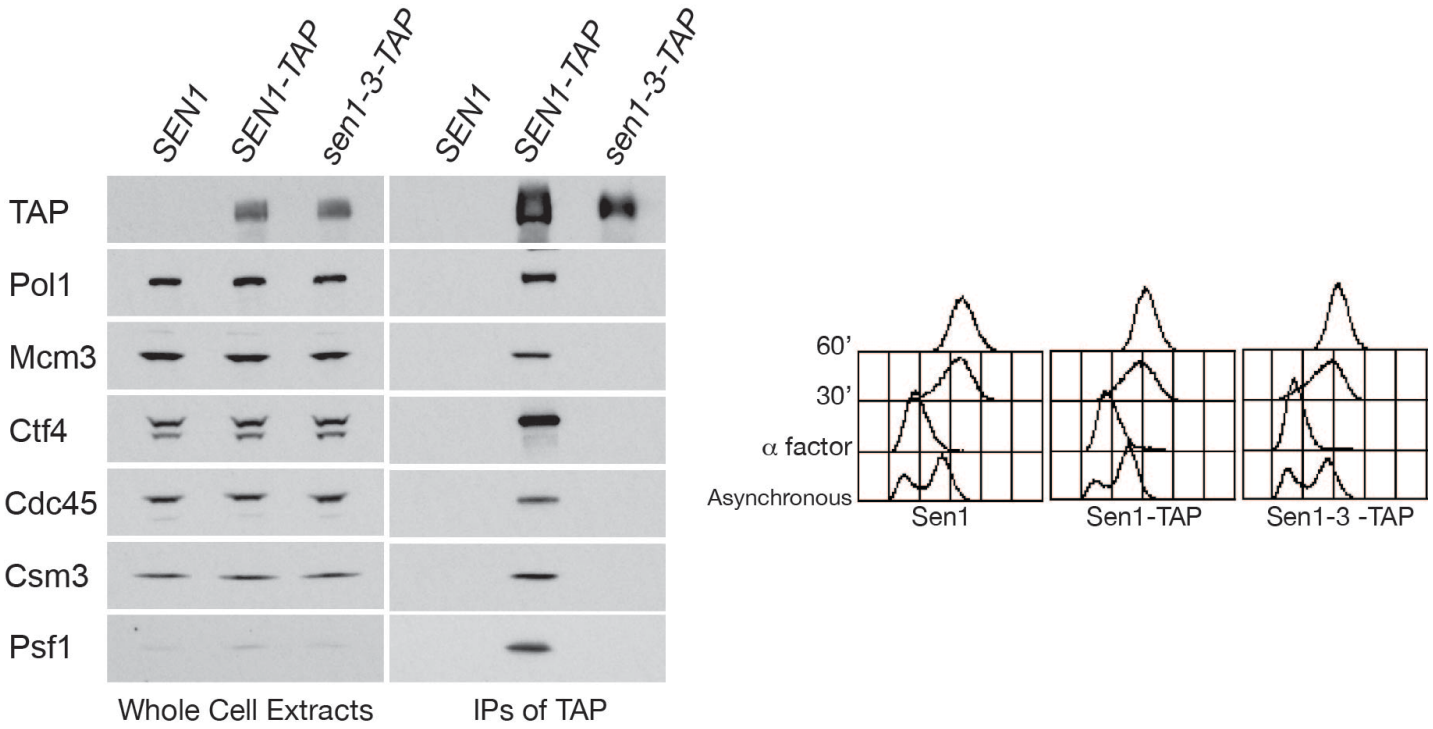


Figure 4.31. The *sen1-3* allele abrogates the interaction between Sen1 and the rest of the replisome when expressed at the genomic locus. Yeast cells carrying either the untagged *SEN1* allele or *SEN1-TAP* or *sen1-3-TAP* were grown in YPD at 24°C to a density of 0.7 x 10^7 cells/ml. The cells were synchronized in G_1 by addition of α-factor to a final concentration of 7.5 µg/ml for 3 h and the α-factor was washed out with fresh YPD and cells were released in S phase at 24°C. 30 min after release, cells were harvested at −80°C and the samples were used to generate cell extracts. The liberated proteins were immuno-precipitated using magnetic beads conjugated to IgG antibodies. The cultures were also harvested for FACS profile when asynchronous, once synchronized in G_1, 30 min after release (at point of harvest) and 60 min after release. WCE and IPs are shown on the left hand and the FACS samples are shown on the right. Strains used: CS**74**, CS**2853**, CS**2854**.

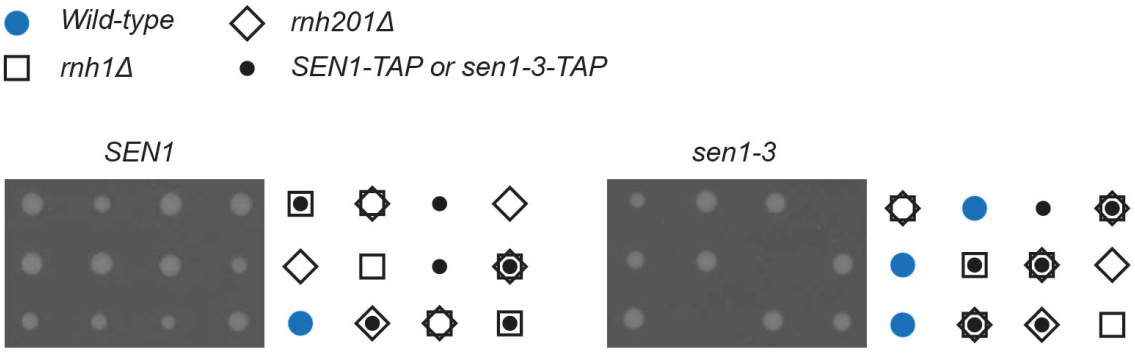


Figure 4.32. The triple *rhn1Δ rhn201Δ sen1-3* mutant is inviable for growth. A strain carrying the *rhn1Δ rhn201Δ* deletion mutations was crossed with strains carrying either the *SEN1-TAP* or *sen1-3-TAP* alleles. The resulting diploids were transferred on sporulation medium and upon formation of asci, the spores within individual asci were aligned by micro-manipulation. The spores were allowed to germinate and grow on rich YPD medium at 24°C before being scored by replica-plating on selective media. Parental strains used: CS**2735**, CS**2853** and CS**2854**

CHAPTER 5: DISCUSSION

5.1 THE ROLE OF POLYMERASE α DURING IMPRINTING AT *MPS1* IN *SZ. POMBE*

Fission yeast relies on the asymmetric nature of DNA replication to imprint and switch cell-types (Dalgaard and Klar, 1999). Efficient imprinting requires unidirectional replication of the *mat1* locus which is itself dependent on forks to terminate at *RTS1* (Dalgaard and Klar, 2001) and to pause at *MPS1* (Dalgaard and Klar, 2000). Several protein factors are required for the imprinting process. Amongst them, Pol1 is unique in that it is not required for the barrier activity at either *MPS1* or *RTS1*. Instead, Pol1 is able to respond to competent fork stalling at *MPS1* to imprint at that locus. Interestingly, a single switching- and imprinting-defective allele of *pol1* (*swi7-1*) has been isolated to date.

Here, I have used gel shifts assays to show that recombinant Swi7-1 has decreased affinity for a DNA/RNA substrate compared to its wildtype counterpart. Perhaps as a direct consequence of this reduced affinity, Swi7-1 was also seen to be a less processive enzyme compared to Pol1. Crystallographic data of Pol1 suggests that the Gly1116 residue (Ser1134 in *S. cerevisiae*) that is mutated in *swi7-1* is located in the thumb domain of the polymerase and is required for strengthening the grip of this domain on the RNA portion of a DNA/RNA substrate (Fig 5.1). Substituting the glycine to a glutamic acid increases both the steric hindrance and negative charge at that particular residue and is predicted to weaken the interaction between the enzyme and its substrate (Perera et al., 2013). Given that Swi7-1 carries this G1116E substitution, our data fits this prediction.

Figure 5.1. Crystallographic data predicts that the *swi7-1* mutant has reduced affinity for its DNA/RNA substrates because of the G1116E substitution. (A) The Pol1 protein from *Sz. pombe* was aligned to its orthologues from *S. cerevisiae*, *H. sapiens*, *X. laevis* and *C. elegans* using Clustal Omega and the image was generated using T-Coffee (Li et al., 2015). The alignment scores are given in terms of colour; blue/purple denotes residues that are not conserved whilst red/pink marks residues that are conserved. Additionally, the strength of conservation is highlighted from strongly conserved (.), to more conserved (:), to invariant (*). The residue mutated in *swi7-1* (and the corresponding residues in the orthologues of the protein) is highlighted. **(B)** The Ser1134 residue in Pol1$^{S. cerevisiae}$ (that aligns with Gly1116 in Pol1$^{Sz. pombe}$) (highlighted) interacts with the phosphate backbone of the DNA/RNA hybrid (Perera et al., 2013).

This suggests that the efficiency of imprinting correlates with the ability of Pol1 to bind to its nucleic acid substrate at *MPS1*. According to such a model, replacing the incriminating Gly1116 residue with an aspartic acid would result in a mutant that has intermediate affinity for DNA/RNA substrates between those of Pol1 and Swi7-1. In addition, this Pol1G1116D mutant would also have imprinting and cell-type switching efficiencies intermediate between those of Pol1 and Swi7-1. We constructed a novel allele of *pol1* (*swi7-2*) that is characterized by this G1116D substitution and found that cells carrying it indeed had cell-type switching defects intermediate to those of *pol1* and *swi7-1*. This result would suggest that the affinity of Pol1 (and consequently of Polymerase α) for its nucleic acid substrate at *MPS1* is important for imprinting. However, given that neither the binding affinity of Swi7-2 for its substrate nor its polymerase activity has been tested empirically, we cannot at this time exclude the possibility that the defect in the *swi7-2* cells may not be linked to a reduced affinity of Pol1 for its substrates.

The imprint itself is RNase-sensitive and has been shown to consists of up to two ribonucleotides (Vengrova and Dalgaard, 2004, Vengrova and Dalgaard, 2006). Some controversy surrounds the nature of the imprint as it has alternatively been characterized as a nick (Kaykov and Arcangioli, 2004). However, data from both groups are compatible with the imprint being RNA in nature (Vengrova and Dalgaard, 2005). In this study, we have identified a primase allele, *spp1-GFP*, that is completely defective for cell-type switching. We have also shown that *mcl1Δ* mutants are also defective for cell-type switching. However, the switching-defective pattern of *mcl1Δ* mutants is similar to that of *pol1-1* and not to that of *swi7-1*, suggesting a stochastic loss of imprinting in some cells. Mcl1 is known to tether Pol1 to the replisome. Taken together, the characterization of the switching-defects of these mutants, especially of the *spp1-GFP* allele, further strengthens the case for the imprint to be RNA in nature. It should be noted however that iodine staining assesses the ability of cells to switch cell-types and not their imprinting directly. As such, the ability of cells carrying the *swi7-2*, *spp1-GFP* and *mcl1Δ* alleles to imprint need to be assessed formally by Southern blotting. Furthermore, the fact that cells carrying *mcl1Δ* from two sources have non-identical staining phenotypes

need to be reconciled. Importantly, the presence of genetic interactors need to be assessed.

From the literature, it is clear that imprinting is regulated both spatially and temporally. Indeed, at the *mat1* locus, three distinct priming sites have been identified. Two priming sites exist 30 and 350 bases past the *MPS1* locus respectively (Vengrova and Dalgaard, 2004) and a priming site occurs at *MPS1* itself (Sayrac et al., 2011). This unusual arrangement of priming sites is thought to contribute to the formation of the imprint at *MPS1* by using an **RNA primer as a precursor**. In addition, Pol1 is required for the imprinting process, downstream of a fork pausing event at *MPS1*.

Our results suggest that Pol1 participates in imprinting in one of two ways. It is possible that Pol1 participates in imprint formation by recruiting factors necessary for processing the RNA primer at *MPS1*. This would need to happen during pausing so that Pol δ does not remove the DNA/RNA primer synthesized by Pol α. In fact, it is possible that the pausing event at the *MPS1* locus enables remodelling of a part of the *mat1* chromatin that includes the 5'-end of the RNA primer into heterochromatin. This would prevent both the synthesis of the *mat1* locus by Pol δ and removal of the RNA primer originating from *MPS1* (Devbhandari et al., 2017), allowing time for the primer to be processed into the imprint. Given that Swi7-1 has reduced affinity for its DNA/RNA substrate, it might disengage from the substrate too quickly to recruit factors that are required to process the primer into the imprint (Fig 5.2). This would lead to temporal dysregulation of the imprinting process.

Another scenario reflects the fact that the Swi7-1 is defective in polymerizing nucleotides *in vitro*. Here, the primase subunit of Polymerase α still synthesizes a primer that is converted post-replicatively into the imprint in wildtype cells. The primer is then converted in a Polymerase α- independent manner into the imprint. However, because of the reduced affinity of Swi7-1 for its substrates, it is possible that upon intramolecular hand-off of the substrate from $Spp1^{Pri1}$ to Swi7-1, the latter can only synthesize DNA inefficiently so that $Spp1^{Pri1}$ needs to **prime more frequently** as a

consequence. This would lead to atypical primers that are either inappropriate precursors for formation of the imprint or that are efficiently removed by Pol δ (Devbhandari et al., 2017). An alternative to atypical priming in *swi7-1* cells could be **abortive priming**, where Swi7-1 can only synthesize a few bases before falling off its substrate. The DNA/RNA substrate are not processed quickly enough so that RNases (or Pol δ) might remove the ribonucleotides in these hybrids. This antagonizes the formation of the imprint at *MPS1* (Fig 5.3). Both high frequency and abortive priming can be thought as spatial dysregulation of imprinting.

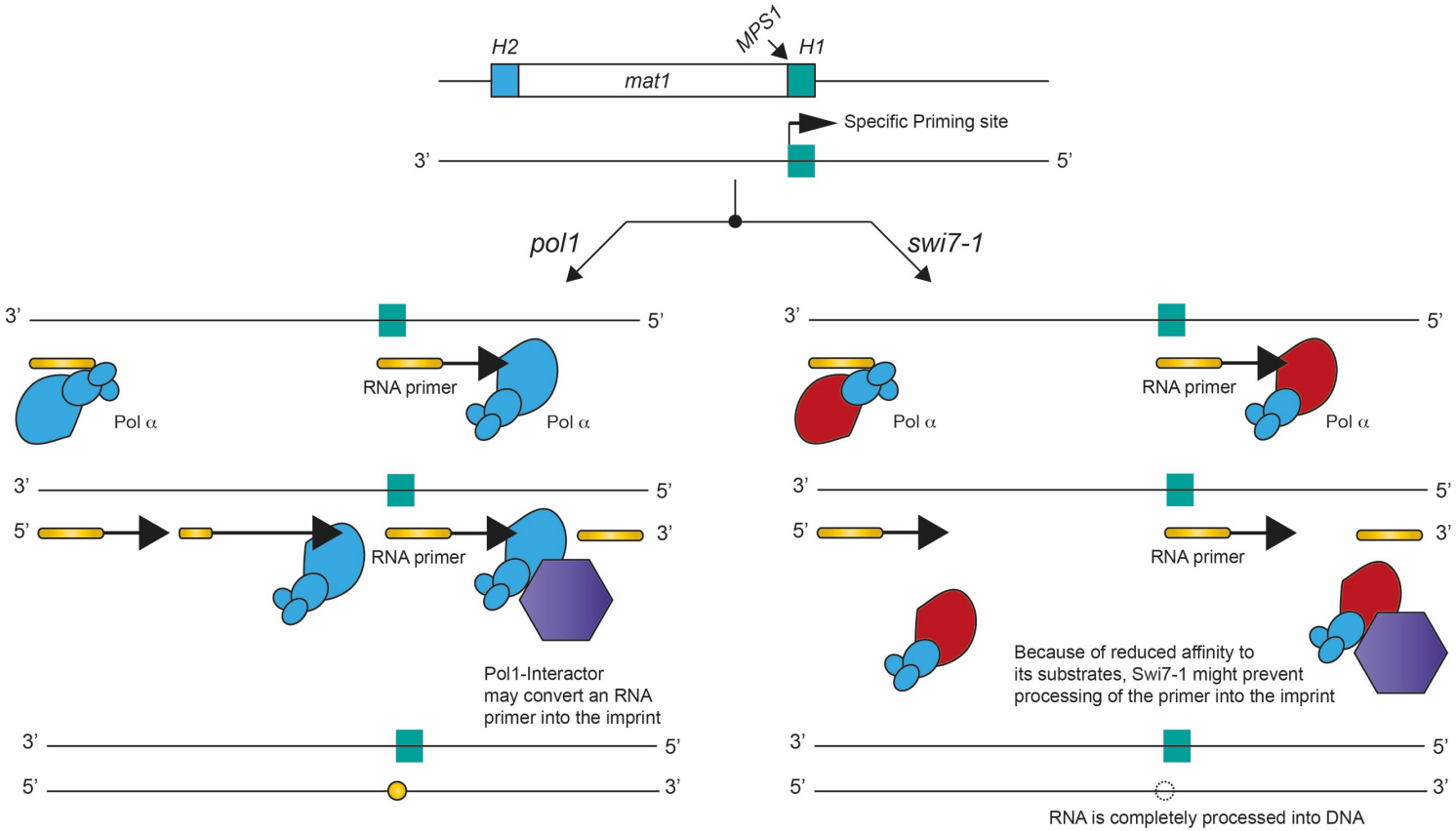


Figure 5.2. Pol1 could participate in maturation of the imprint from a precursor RNA primer. In wildtype cells, the imprint could be generated from an RNA primer at *MPS1*. Pol1 could be important in this process by recruiting a protein that processes the primer into the imprint at the junction of the *mat1* cassette and its *H1* locus. By contrast, the Swi7-1 mutant has reduced affinity for its substrates and this would preclude formation of the imprint.

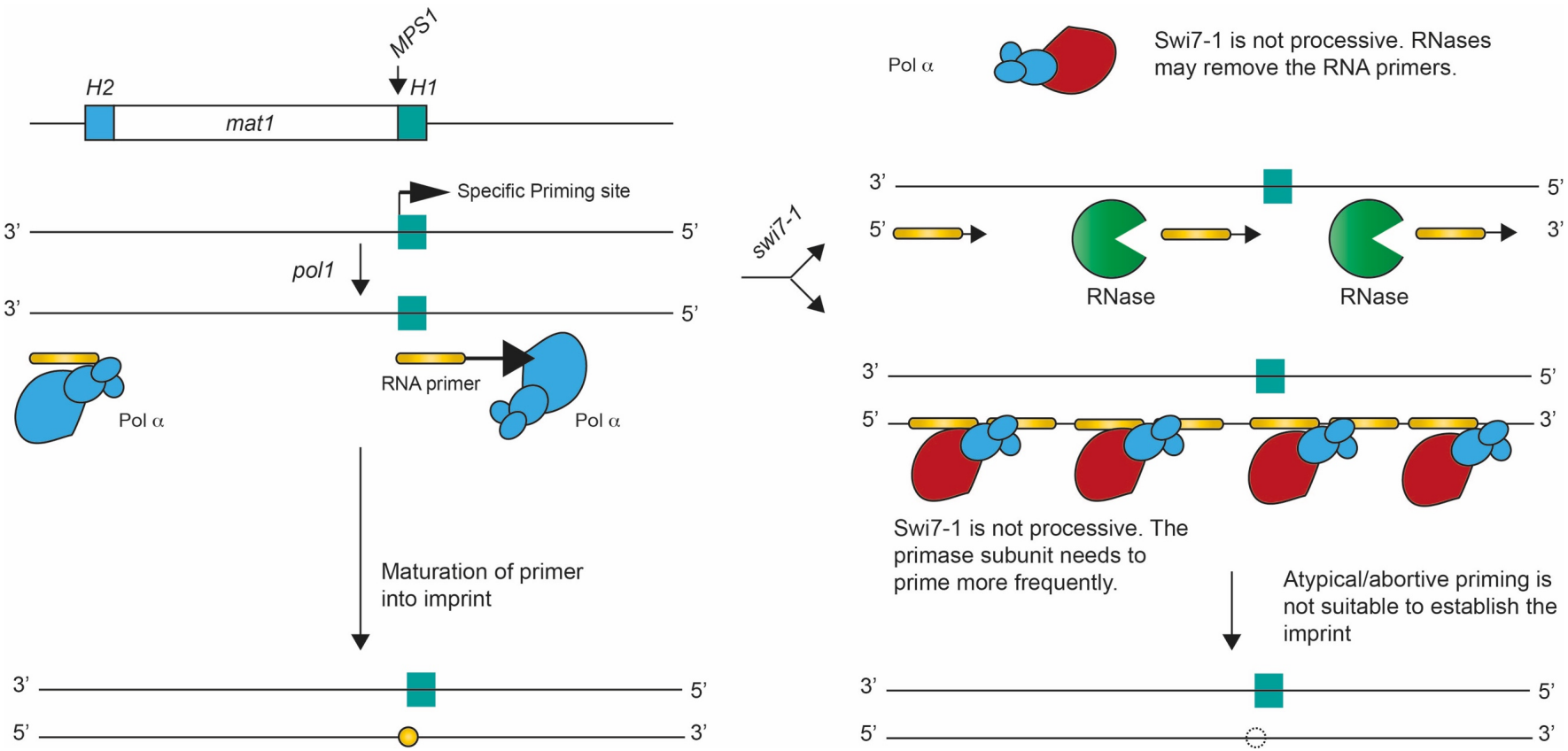


Figure 5.3. Proper synthesis of DNA/RNA hybrids by Polymerase α could be important for imprinting at *MPS1*. In wildtype cells, proper synthesis of the DNA/RNA hybrid could enable processing of the RNA primer at *MPS1* into the imprint. However, it is possible that Swi7-1 is not a good polymerase *in vivo*. To compensate, the primase subunit needs to synthesize more frequently. These atypical primers are not suitable precursors for imprint formation. Alternatively, once primase hands the primer over, inefficient activity by Swi7-1 may eventually lead RNases to remove the RNA primer prematurely so that the imprint is not synthesized.

5.2 EXPLOITING THE *SWI7-1* MUTANT AS A TOOL TO STUDY RESPONSES TO FORK BARRIERS AND TO UNDERSTAND ORIGIN USAGE

Here, I have demonstrated that Swi7-1 has reduced affinity for its template. Whether this also occurs *in vivo* has not been addressed. To do so, chromatin binding assays (Donovan et al., 1997, Liang and Stillman, 1997) could be used whereby spheroplasted cells are lysed and chromatin pellets and soluble fractions are collected separately. The relative binding of Pol1 and Swi7-1 to chromatin could thus be compared. Alternatively, the relative affinities of these two variants for nucleic acid substrates could be assessed by chromatin-immuno-precipitation. In fact, these experiments would be critical to determine whether the decreased affinity of Swi7-1 for its substrates is compensated *in vivo* by other components of the replisome, such as Mcl1 (Mcl1 could increase the local concentration of the enzyme at forks). Should Swi7-1 also have defects in binding to nucleic acids *in vivo*, this would be in agreement with the model of imprinting proposed here. In addition, this would open new avenues for research.

For instance, if *swi7-1* cells are characterized by abortive priming, this could be a useful tool to study origin usage. Unlike origins in *S. cerevisiae* that contain an 11-17 base consensus sequence (Theis and Newlon, 1997), origins in *Sz. pombe* lack consensus, although they are AT-rich. However, both the location and firing efficiency of these origins have been systematically characterized (Heichinger et al., 2006). Thus, systematic analysis of origin firing in *swi7-1* cells could inform us about the protein requirement for efficient origin firing in an *in vivo* setting. In this highly speculative scenario, it would also be interesting to determine whether some origins are more or less refractory to the *swi7-1* mutation and whether this putative feature would be sequence specific.

Should *swi7-1* cells be characterized by high frequency priming instead, it can be hypothesized that such priming would not be restricted to the vicinity of the

MPS1 locus. It is also known that stalled forks can eventually skip barriers by re-priming immediately past them (Elvers et al., 2011, Lopes et al., 2006, Yeeles and Marians, 2013). It would be interesting to see if excessive priming can confuse replisomes, thus effecting a poor response to fork stalling.

5.3 REASSESSING THE POSTTRANSCRIPTIONAL CONTROL OF *SEN1* EXPRESSION

The earliest attempt to characterize the enzymatic function of Sen1 *in vitro* was made in the Hochstrasser lab where the full-length gene (*SEN1*) was cloned in a high-copy vector plasmid and transformed in an *S. cerevisiae* host (DeMarini et al., 1995). For comparison, the authors also cloned a truncated allele (*sen1-2*) in the high-copy plasmid that they transformed in an *S. cerevisiae* host. This *sen1-2* allele encodes for the Sen1 (976-2231) protein. However, whilst the Sen1-2 protein was highly visible on Western blots, the wildtype protein was not (DeMarini et al., 1995). On the other hand, Northern blots showed that the two genes were expressed at comparable levels (DeMarini et al., 1995). The authors also cloned *SEN1* constructs fused to lacZ from *E. coli* in a high-copy vector and found that both Sen1 (1-1742)-LacZ and Sen1 (433-1742)-LacZ were expressed poorly whilst both Sen1 (571-1742)-LacZ and Sen1 (857-1742)-LacZ were highly expressed. This suggested that residues 434- 570 of Sen1 negatively auto-regulated its protein levels by post-transcriptional control of its transcript.

In the present study, the *TAP-SEN1 (2-931)*, *TAP-SEN (2-1103)* and *TAP-SEN1 (2-1901)* constructs were expressed under control of the strong *GAL1* promoter. Although the mRNA levels of these different constructs were not formally assessed, appreciable amounts of proteins were recovered. This suggests that residues 434-570 of Sen1 are required but are insufficient to dampen the protein levels of Sen1 by post-transcriptional control of *SEN1* transcripts. Instead it suggests one of two likely models. Either a second stretch of residues within the Sen1 protein co-operatively mediates post-transcriptional control of its transcript. Alternatively, residues 434-570 alone are involved in this post-transcriptional control but the target within the *SEN1* transcript is located beyond the first 3309 bases. Given Sen1's ability to moderate its expression by affecting gene dosage dependence, it would be interesting to investigate whether Sen1 also affects the regulation of transcription at other loci.

Here, we generated a yeast strain with the following genotype: *MATα td-MYC-sen1-1* (*klTRP1*+) *GAL1-UBR1* (*HISMX*) *GAL1-TAP-SEN1 (2-2231)* (*LEU2*+) (CS2188). The isogenic *MATa* strain was first created as a control where Sen1 could be depleted at non-permissive temperatures and replaced with TAP-Sen1 (2-2231). However, the TAP-Sen1 (2-2231) was seen to suppress *sen1-1* lethality at 37°C on YPD. This was seen in all clones with this genotype (the *GAL1-TAP-SEN1 (2-2231)* was fully sequenced and independently cloned at the *LEU2* locus thrice). Given that the marker for both *td-MYC-sen1-1* and *GAL1-TAP-SEN1 (2-2231)* (both markers were present C-terminally to their respective gene) were not compromised, it is unlikely that the leaky expression of *GAL1-TAP-SEN1 (2-2231)* arose as a result of meiotic recombination between those two alleles to generate a wildtype *SEN1* gene. Instead, this observation suggests that the *td-MYC-sen1-1* allele is defective in repression of the *GAL1* promoter. The fact that the other *SEN1* constructs did not suppress the lethality of *td-MYC-sen1-1* at 37°C on YPD could indicate that the other constructs are not as well suited to complement Sen1's depletion at low concentrations.

5.4 STABILITY OF SEN1 CONSTRUCTS

When overexpressed, Sen1 (2-931), Sen1 (2-1103) and Sen1 (2-1901) were easily detected in both TCA protein samples and cell extracts upon rupturing of the cell walls and membranes. This contrasted with Sen1 (2-2231), Sen1 (931-2231) and Sen1 (1095-2231). These constructs were detected in TCA samples but underwent rapid lysis in cell extracts. In fact, when using 1.75 x 10^9 cells to generate the extracts, these constructs could not be detected by Western blotting (Fig 5.4). This was independent of the protein tag used. This observation suggests that the last 330 residues of Sen1 contain a degron that can be targeted for rapid proteolysis. Constructs carrying these 330 residues were only visible in cell extracts prepared from 7 x 10^9 cells. This would suggest that the protein levels of Sen1 are maintained beneath a certain threshold, both by posttranscriptional control of its mRNA (DeMarini et al., 1995) and by rapid proteolysis. The implication is evident. Although Sen1 is necessary for viability, excessive levels of the protein would cause global problems in transcription termination and gene expression, affecting growth.

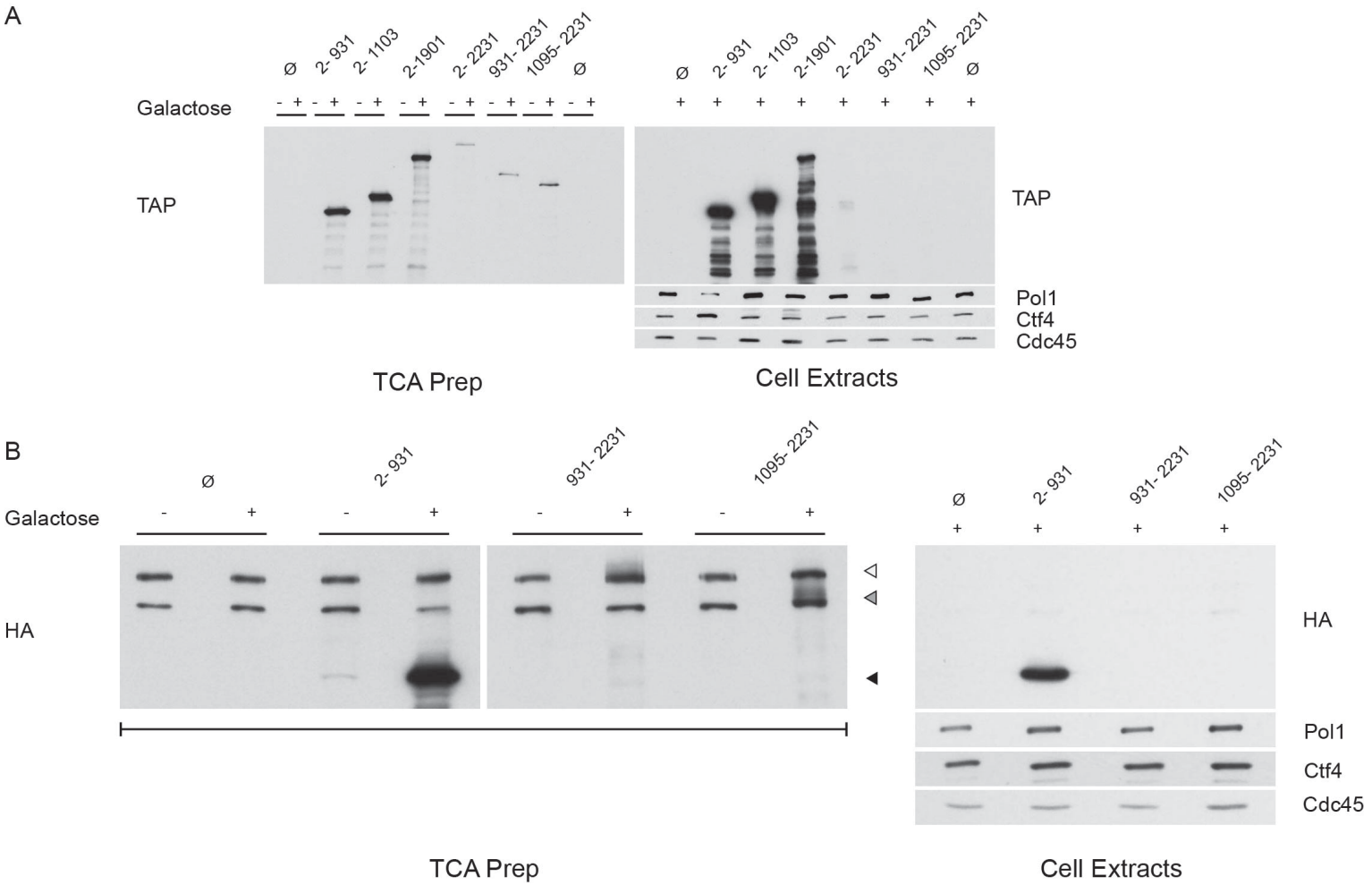

A
Ø
2- 931
2- 1103
2-1901
2- 2231
931- 2231
1095- 2231
Ø
Galactose
- + - + - + - + - + - + - +
TAP
TAP
Pol1
Ctf4
Cdc45
TCA Prep
Cell Extracts
B
Ø
2- 931
931- 2231
1095- 2231
Galactose
- + - + - + - +
HA
HA
Pol1
Ctf4
Cdc45
TCA Prep
Cell Extracts

Figure 5.4. Constructs of Sen1 encoding the last 330 residues were especially liable. Sen1 constructs were cloned under the *GAL1* promoter. To assess expression after addition of galactose, cells were collected for protein extraction by TCA. Cells were also collected to make cell extracts. Comparison between TCA samples and cell extracts showed that constructs of Sen1 carrying the last 330 residues were very liable. By contrast, constructs carrying the first 1901 residues and endogenous proteins were relatively stable. This suggests that the last 300 residues of Sen1 encode for a degron that cells use to keep Sen1 levels beneath a certain threshold. **(A)** Constructs tagged with the TAP tag collected in S phase. **(B)** Constructs tagged with 3HA collected in S phase. The arrows show the positioning of the different Sen1 constructs [black arrow: Sen1 (2- 931), light grey arrow: Sen1 (931-2231) and dark grey arrow: Sen1 (1095-2231)].

5.5 SEN1 INTERACTS PHYSICALLY WITH NRD1, RAD5 AND RNA POLYMERASES

In order to identify novel interactors of Sen1, a Y2H screen was employed with Sen1 (2- 931) used as bait. From the screen, two novel **physical** interactors of Sen1 were identified. The first one was Nrd1. It has previously been shown that Nab3, Nrd1 and Sen1 form a complex, and that Sen1, via residues 1890-2092, makes direct physical contact with Nab3 (Nedea et al., 2008). Here, the first 282 residues of Nrd1 was sufficient for the interaction with Sen1 (2-931), at least in the Y2H screen. If this interaction can be verified biochemically, it would suggest that the individual components of the NNS complex can form direct physical contacts with one another. This would also suggest that this trimer can toggle between an open and a closed conformation. This may have interesting implications for our understanding of the structure and assembly of the complex. This putative ability to open or close could also be important functionally (Fig 5.5).

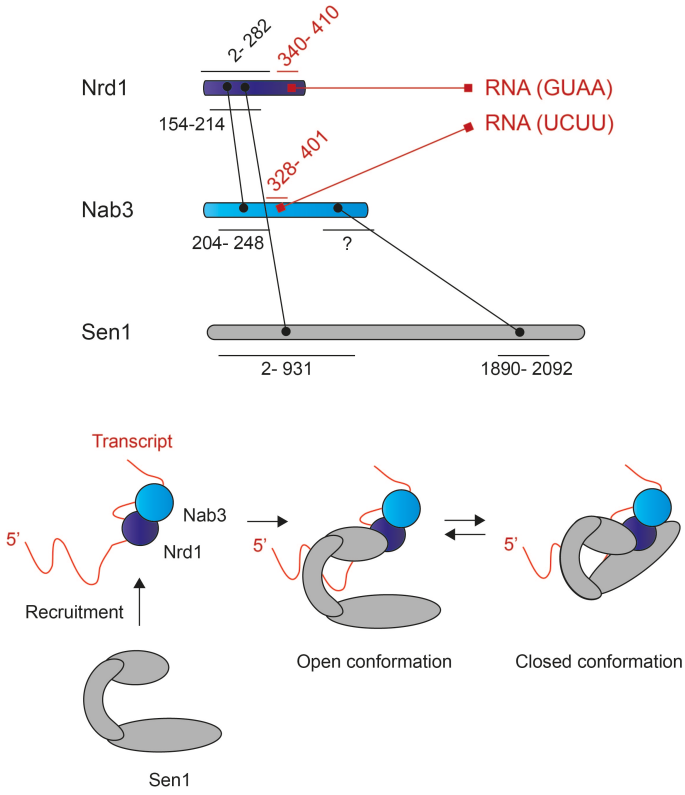


Figure 5.5. Identification of Nrd1 as a direct interactor of Sen1 suggests that the NNS complex can toggle between an open and a closed conformation.

It has previously been shown that Sen1 interacts physically with Rad2 (Ursic et al., 2004), a protein involved in nucleotide excision repair (Ursic et al., 2004). NER is one of the pathways that cells use to remove pyrimidine dimers upon UV-mediated damage (Mao et al., 2016) and Sen1 was later found to play an important role in NER itself by virtue of its N-terminal domain (Li et al., 2016). Importantly, the triple *rad7*Δ *rad26*Δ *SEN1(1089-1929)* mutant was more susceptible to UV treatment than the double *rad7*Δ *rad26*Δ mutant. Deletion of both *RAD7* and *RAD26* prevents repair of UV-mediated lesions using either branch (global genomic NER and transcription-coupled repair) of the nucleotide excision repair. Taken together, these observations suggest that Sen1 participates in the removal of UV-induced lesions in an NER-independent pathway.

One alternative way through which UV damage can be dealt with is the DNA-damage tolerance (DDT) pathway where lesions can simply be skipped during DNA replication (Xu et al., 2015). Within DDT, a sub-pathway known as error-free DDT is thought to use the newly replicated 'sister' strand as an undamaged template to bypass the damage site (Xu and Clayton, 1996). This is not dissimilar to replication between the *RTS1* and *MPS1* loci in imprinted fission yeast strains **lacking the donor loci** (Klar and Miglio, 1986). Error-free DDT requires the poly-ubiquitination of PCNA that is mediated by three proteins: Rad5, Mms2 and Ubc13. Rad5 is the E3 ligase. Here, in the Y2H assay, Rad5 has been identified as a physical interactor of Sen1. Notably, a Rad5 (12-265) construct was sufficient for this interaction. Taken together with observations in *rad7*Δ *rad26*Δ *SEN1(1089-1929)* cells, this suggests that Sen1 may be involved in error-free DTT by virtue of its N-terminal domain.

It should be noted that strains used in this study were all isogenic to W303 that carries the *rad5-535* mutation. In the W303 background, cells carrying the *sen1-3* mutation did not show any defect compared to *SEN1* cells. Moreover, there was no observable difference between cells with the *sen1-3 RAD5* and *sen1-3 rad5-535* genotypes (not shown). Thus, the *sen1-3* and *rad5-535* mutants do not interact genetically. Identifying alleles of *SEN1* and *RAD5* that would interact genetically with one another would be useful to discern the

biological relevance of Sen1's interaction with Rad5. This constitutes an interesting avenue for future research.

By mass spectrometric analysis of proteins that co-precipitate with Sen1 (2-931), the N-terminal domain of Sen1 was also shown to interact with RNA Pol I, II and Pol III. RNA Pol I and III lack a CTD (Wild and Cramer, 2012), suggesting that the interaction between Sen1 (and the NNS complex) and these two RNA polymerases is different to the interaction between the NNS complex and RNA Pol II. Importantly, the number of peptides detected for the two largest subunits of RNA Pol III in the mass spectrometric screen was at least 10-fold higher than the number of peptides for the corresponding subunits of RNA Pol I and II. This suggests that Sen1 has a preference for RNA Pol III and that the protein has a critical role in the homeostasis of transfer RNA. Consequently, Sen1 is involved in regulation of the translation process.

No replisome components were identified with the Y2H screen. This could be a result of either the technical limitations of the screen or of the physiological aspects of Sen1's interaction with the replisome or a combination of the two. For example, should the trimerization of Ctf4 be necessary for interaction with Sen1, it would be possible that the Ctf4 fragments used in the Y2H screen were not able to trimerize or, alternatively, trimerized in such a way as to prevent interaction with the Sen1 construct. It should be noted however that Rad5 is required for fork progression following MMS treatment. Moreover, Rad5 also forms subnuclear foci upon MMS treatment and its protein levels peak in S phase (Ortiz-Bazán et al., 2014). Consequently, whilst not strictly involved in DNA replication, Rad5 is nonetheless required for fork progression under certain conditions.

5.6 THE DNA/RNA HELICASE SEN1 IS A *BONA FIDE* COMPONENT OF THE REPLISOME IN THE MODEL ORGANISM *S. CEREVISIAE*

It has previously been shown that Sen1 travels with forks in *S. cerevisiae* (Alzu et al., 2012). In fact, depletion of Sen1 by shifting cells containing the temperature-sensitive *sen1-1* allele to non-permissive temperatures leads to a lengthening of S phase and terminal arrest in G_2. These observations hint at a role of Sen1 during DNA replication. Here, using IPs from synchronous cultures, I have shown that Sen1 is a component of the replisome in yeast. It is perhaps surprising that Sen1 had yet to be recognized as such. Although no screen can claim to be exhaustive, one plausible explanation for this delay is the liability of the protein (DeMarini et al., 1995). Indeed, in this study IPs from cell extracts of 1.75×10^9 cells did not reveal any interaction between Sen1 and other replisome components, although this is sufficient to show interaction amongst other replisome components. It is only when I raised the yield to 7×10^9 cells that physical interaction between Sen1 and replisome components could be visualized. Importantly, full-length Sen1 would not seem to interact with any component of the replisome outside of S phase.

5.7 POSTTRANSLATIONAL MODIFICATIONS OF SEN1 OR ITS BINDING PARTNERS MAY AFFECT THE INTERACTION WITH THE REPLISOME

On Western blots, Sen1 is often seen as a smear rather than a discrete band. This is more evident in immuno-precipitated samples. Sen1 is a phosphoprotein and residues spanning positions 2125 to 2144 (ILTASDYGEPNQNGQNGANR) were identified in a mass spectrometric screen of the phosphoproteome of *S. cerevisiae* (Bodenmiller et al., 2010). By employing similar proteomics techniques, several other residues have been shown to be phosphorylated (Table 5.1).

Table 5.1. Phosphorylated sites within Sen1

Position of Residue	Reference
S863	(Swaney et al., 2013)
S1047	(Swaney et al., 2013)
S1053	(Swaney et al., 2013)
S1055	(Swaney et al., 2013)
S1058	(Swaney et al., 2013)
T1487	(Bodenmiller et al., 2010)
S1505	(Swaney et al., 2013)
S2218	(Helbig et al., 2010)

It is as yet unknown whether phosphorylation is a requirement for Sen1 to interact with the rest of the replisome. Our immuno-precipitation assays include two phosphatase inhibitors: sodium fluoride (that inhibits acidic, phosphoseryl and phosphothreonyl phosphatases) and sodium β-glycerophosphate (that inhibits phosphoseryl and phosphothreonyl phosphatases). It would be relatively straightforward to assess whether phosphorylation is important for the interaction between Sen1 and other replisome components. To do so, a two-step S phase IP of TAP-Sen1 can be employed where, after cleaving Sen1 from TAP beads, the elute would then

be split and treated with either an inactivated or a functional phosphatase prior to incubation with calmodulin beads. However, it would not be possible to distinguish whether phosphorylation of Sen1 or the phosphorylation of some other replisome component (such as Ctf4) is important for the interaction using this method.

Interestingly, I have shown that full-length Sen1 does not interact with any replisome components outside of S phase. However, a truncated variant of the protein, Sen1 (2-931) is able to interact with Ctf4 and GINS throughout the cell cycle. Absence of the first 931 residues abrogates the interaction with the replisome in S phase, suggesting that the last 1300 residues of the protein are dispensable for interaction with the replisome. The difference between Sen1 and Sen1 (2-931) suggests that the last 1300 residues regulate the interaction between Sen1 (2-931) and Ctf4 and GINS in a cell cycle-dependent manner. This regulatory role could be dependent on posttranslational modification(s). Alternatively, a cell cycle dependent interactor of Sen1 that antagonizes its interaction with replisome components may require some post-translational modification as a cue to break its interaction with Sen1.

5.8 IDENTIFYING REPLISOME INTERACTORS OF SEN1

By immunoprecipitation, the Sen1 (2-931) construct was seen to interact with both Ctf4 and GINS throughout the cell cycle. Deletion of Ctf4 abrogates interaction with GINS, suggesting that Sen1 interacts with Ctf4 only and not with GINS. However, it is also possible that Sen1 straddles both Ctf4 and GINS. To determine whether this is the case, GINS can be degraded specifically by shifting cells carrying the temperature-sensitive *psf1-1* allele to non-permissive temperatures. IPs in the absence of Psf1 can then be carried out. Should Sen1 (2-931) not interact with Ctf4 in the absence of Psf1, this would suggest that Sen1 requires both Ctf4 and GINS to interact with the rest of the replisome. Moreover, the results presented here do not exclude the possibility that Ctf4 interacts with Sen1 through some other unknown interactor(s). To show that Ctf4 and Sen1 directly interact with one another, the two proteins need to be expressed and purified in isolation (for example, in an *E. coli* host or in a cell-free translation system) and the two proteins can then be mixed. Formation of Sen1-Ctf4 complexes would conclusively support the ability the two proteins to directly interact with one another.

Ctf4 is made of three domains: an N-terminal WD40 domain, a central β-propeller domain (required for trimerization) and a C-terminal α-helical domain that houses a second WD40 region. The first 350 residues of Ctf4 on their own cannot interact with Sen1 whilst deletion of the first 383 residues does not affect the interaction with Sen1. This suggests that the N-terminal domain of Ctf4 is dispensable for its interaction with Sen1. Truncation of the α-helical domain leads to a pronounced reduction in the levels of Ctf4 binding to Sen1 (2-931), suggesting that this domain is required for optimal interaction. However, given that Sen1 (2-931) still interacts with Ctf4 (1-841), this suggests that either Sen1 (2-931) binds to at least two different regions of Ctf4 or that Ctf4 trimerization is sufficient for the Sen1 to interact with the protein.

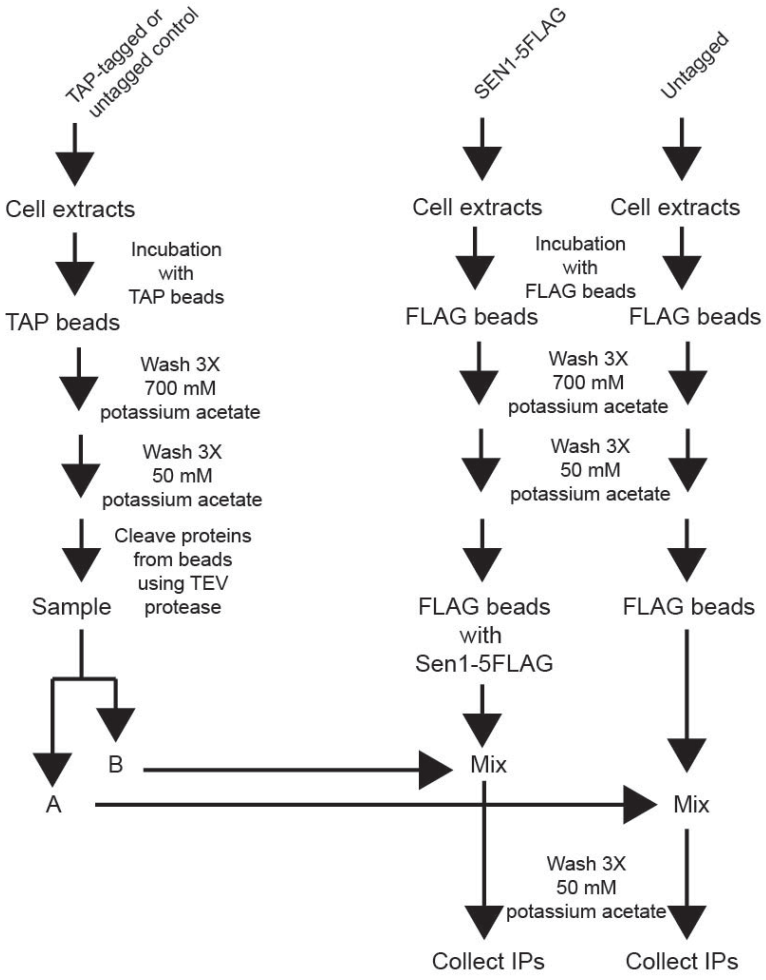
In cells deleted for Ctf4, subpopulations of Sen1 and Sen1 (2-931) can still interact with other components of the replisome in S phase. This suggests that

Sen1 interacts with replisome components other than Ctf4 and GINS. An alternative explanation is that *ctf4*Δ can lead to various changes in the cell (including changes in protein levels) that could contribute to Sen1 being able to artificially bind to the replisome in the absence of Ctf4. One way to test this hypothesis is to substitute a mutant of Ctf4, *ctf4-4E* [*ctf4 (L867E A871E A897E I901E)*] (Villa et al., 2016) for *ctf4*Δ in Sen1 or Sen1 (2-931) IPs. The *ctf4-4E* mutation abrogates the interaction between CIP-containing proteins (such as Pol1 and Sld5) and Ctf4 but it lacks many of the phenotypes of *ctf4*Δ.

Besides Ctf4 and GINS, no other replisome components co-precipitated with Sen1 (2-931) outside of S phase. Either there is an additional regulatory mechanism that prevents this Sen1 construct from binding to its other putative partners outside of S phase or there is a cell cycle dependent restriction to such interactions. As such, to test whether Sen1 interacted directly with other components of the replisome, I have used an *in vitro* reconstitution assay. In brief, strains carrying either the *TAP-MCM3*, *POL12-TAP*, *DPB2-TAP* alleles or an untagged control strain were harvested in S phase. In parallel, a strain carrying the *SEN1-5FLAG* allele and another untagged control strain were also harvested in S phase. Cell extracts were generated for the TAP-tagged strains (and the control) and a modified two-step TAP IP was carried out. After incubation with cell extracts, the TAP beads were washed successively with a solution containing 700 mM potassium acetate to remove weak interactors and a solution containing 50 mM potassium acetate, followed by cleavage of the proteins off the beads using the AcTEV protease. Simultaneously, cell extracts were generated using the cells carrying the *SEN1-5FLAG* allele (and the control) and FLAG IPs were carried out. The FLAG beads were washed successively with a solution of 700 mM potassium acetate and a solution of 50 mM potassium acetate. The CBP-tagged proteins (and their strong interactors) cleaved from TAP beads were then split in half. One half was incubated with FLAG beads to which SEN1-5FLAG were bound whilst the other half was incubated with FLAG beads incubated previously in cell extracts with no tagged protein. Unfortunately, results from this *in vitro* assay were not fully reproducible and the assay needs to be optimized further.

However, if results from this assay were to be correct (Fig 5.6), it would show that Sen1 is able to bind independently to several replisome components. This would not fit with observations that Sen1 (2-931) cannot interact with replisome components in S phase (other than Ctf4 and GINS) upon degradation of Sld3-7-TD in G_1. (Fig 4.20). There are a couple of plausible explanations for the discrepancies between the two assays.

Firstly, it is possible that, within an assembled replisome, its individual components are modified in such a way that allows Sen1 to bind to them. When replisomes are not assembled at origins, the necessary modifications are not present so that some replisome components can no longer interact with Sen1. A second possibility is that Sen1 interacts with some replisome components indirectly through histones or nucleosomes. The inability to form replisomes would interfere with the interaction between histones or nucleosomes with replisome components, also interfering with the ability of these components to interact with Sen1. Finally, some interaction between Sen1 and replisome components could reflect nucleic acid-protein interactions. Indeed, although we use a universal nuclease in our IP experiments, some ssDNA could be trapped in polymerases. Sen1, but not Sen1 (2-931), would then be able to bind to these nucleic acid species by virtue of its helicase domain.


TAP-tagged or untagged control
SEN1-5FLAG
Untagged
Cell extracts
Cell extracts
Cell extracts
Incubation with TAP beads
Incubation with FLAG beads
TAP beads
FLAG beads
FLAG beads
Wash 3X 700 mM potassium acetate
Wash 3X 700 mM potassium acetate
Wash 3X 50 mM potassium acetate
Wash 3X 50 mM potassium acetate
Cleave proteins from beads using TEV protease
Sample
FLAG beads with Sen1-5FLAG
FLAG beads
B
Mix
A
Mix
Wash 3X 50 mM potassium acetate
Collect IPs
Collect IPs

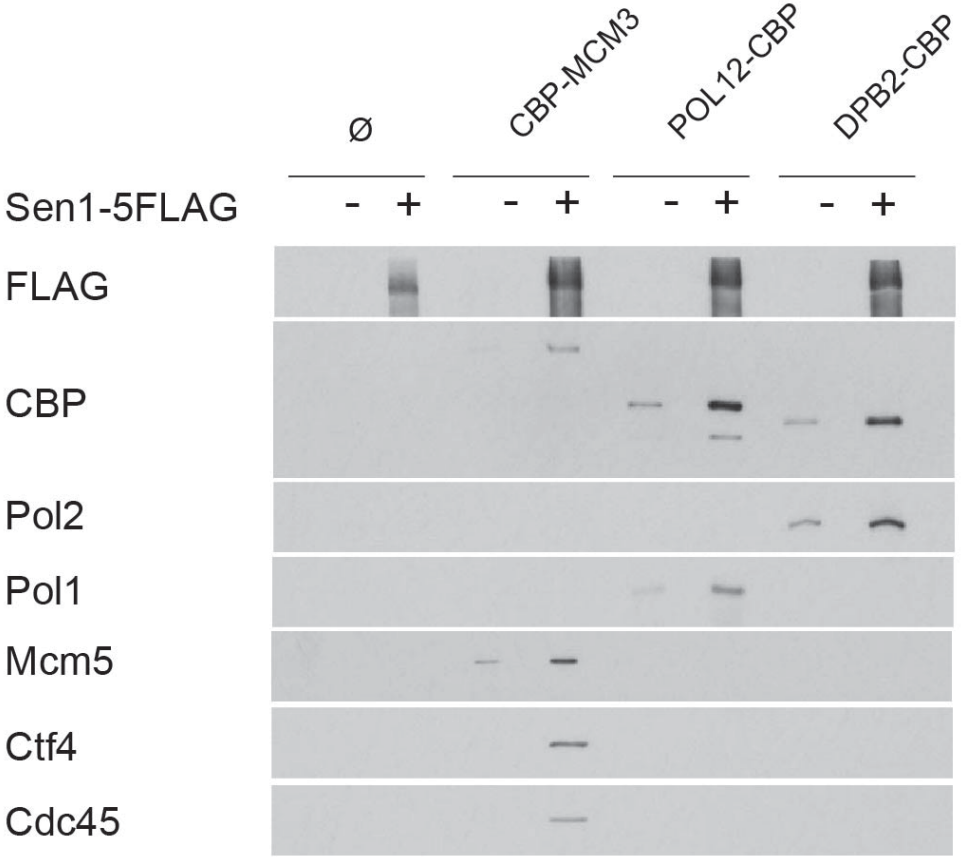

Ø
CBP-MCM3
POL12-CBP
DPB2-CBP
Sen1-5FLAG
- + - + - + - +
FLAG
CBP
Pol2
Pol1
Mcm5
Ctf4
Cdc45

Figure 5.6. Summary of the *in vitro* assay to test phase-specific protein-protein interactions. Cells growing in S phase were harvested. These cells were used to generate cell extracts. Cell extracts from TAP-tagged strains (and an untagged control) were incubated with TAP beads and washed stringently to ensure only TAP-tagged preys and their strong interactors remained on the beads. These proteins were then cleaved from beads using the AcTEV protease. Simultaneously, cells extracts from Sen1-5FLAG strains (and a control) were incubated with FLAG beads. The beads were stringently washed to ensure only FLAG-tagged preys and their strong interactors remained on the beads. Calmodulin-tagged protein samples cleaved from the TAP beads were then split in two and incubated with either the empty FLAG-beads or FLAG-beads bound to Sen1-5FLAG. The left panel gives a schematic of this assay whilst the right panel gives an example output generated. However, results from the screen were not fully reproducible and the screen needs to be optimized further.

5.9 MUTATION OF ARMADILLO REPEAT MOTIFS AS A STRATEGY TO STUDY PROTEIN-PROTEIN INTERACTION

The armadillo (ARM)-repeat is evolutionarily ancient and has evolved to mediate diverse functions (Tewari et al., 2010). In yeast, the Vac8 protein (that shares ~22% identity with the first identified ARM-repeat protein, β-catenin) also has an ARM-repeat motif that serves as a hub for interaction with several proteins. This contributes to the functional versatility of the protein. By performing S phase IPs of MCM2-7 in cells carrying progressively smaller constructs of Sen1, I have identified residues 622-931 of Sen1 that were sufficient for interaction with replisome components. Analysis of these residues revealed that they were predicted to encode several ARM-repeats. Sen1, like Vac8, is a versatile protein with several functions. It is possible that the ARM-repeats contribute to the plethora of activities of the protein. Thus, to create a mutant of Sen1 that no longer interacts with replisome components, I have mutated residues within the ARM-repeats that were predicted to occur on the surface of the secondary structures formed by those repeats and that were highly conserved. Conservation was assessed by comparing the sequence of Sen1 to its close orthologues in other yeasts (Fig 5.7). A single mutant (*sen1-3*) abrogated the interaction between Sen1 and the replisome in S phase. Moreover, several other mutants, including Sen1-4, Sen1-5 and Sen1-6, seemed to interact more strongly with the replisome. These observations suggest that the mutations used effectively altered the conformation of the N-terminal domain of Sen1 without affecting its overall stability. As such the strategy devised to create mutants of Sen1 is a viable one to study protein-protein interactions in ARM-repeat proteins, with the caveat that weakening the interaction with a particular partner could strengthen interaction with other partners and vice-versa. This potential artefact was not tested here but could be assessed by IPs, followed by mass spectrometric analysis.

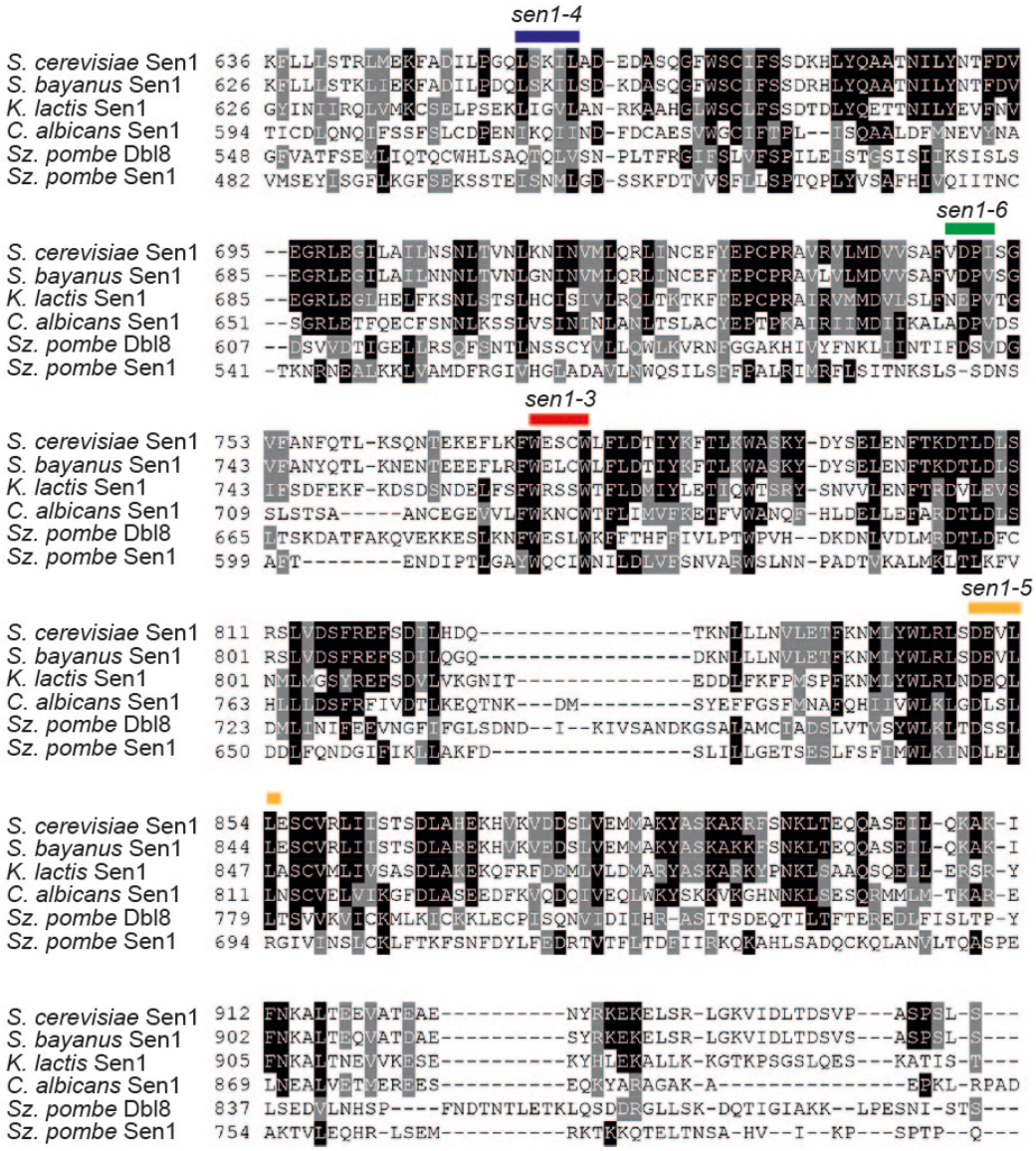


Figure 5.7. Alignment of Sen1 with its close fungal orthologues. In order to generate mutants of Sen1 that do not interact with the replisome, residues were selected based on their conservation status and whether or not they were predicted to be localized on the surface of the secondary structures formed by ARM-repeats. Conservation was determined by comparison with fungal orthologues of the protein. The figure above highlights the residues (*sen1-3*, *sen1-4*, *sen1-5* and *sen1-6*) that were chosen for mutagenesis.

5.10 ASSESSING THE SPECIFICITY OF THE *SEN1-3* ALLELE

Sen1 is diverse in its functioning and it is involved in transcription termination (Mischo et al., 2011) and transcription-coupled repair (Li et al., 2016). The N-terminal domain of the protein is thought to be primarily important for protein-protein interactions and several proteins have been shown to bind to it, including Rad2, Rnt1, as well as RNA Pol II (Ursic et al., 2004, Yüce and West, 2013). In this study, I have identified that Sen1 also interacts with the replisome through this N-terminal domain. To understand the biological relevance of this interaction, it was necessary to generate mutants of Sen1 that lose the interaction with the replisome whilst not interfering with its other interactions and functions. To maximize the likelihood of isolating such a specific mutant, I generated point mutants of *SEN1* and screened them for loss of interaction with the replisome. Using this strategy, the *sen1-3* mutant was isolated. The latter is characterized by three substitutions (W773A E774A W777A).

This novel allele does not present several of the phenotypes associated with the *sen1-2* allele (that encodes for a protein that lacks the N-terminal domain), including slow growth and growth defects on non-fermentable carbon sources such as ethanol and glycerol (Sariki et al., 2016). Nor is it temperature-sensitive, unlike the *sen1-1* allele. Importantly, cells carrying the *sen1-1* allele are also characterized by transcription defects. As such, it would be especially important to assess the transcription termination efficiency in cells carrying the *sen1-3* allele. We are collaborating with Prof Domenico Libri (Institut Jacques Monod, CNRS, France) to assess whether *sen1-3* is characterized by any transcription termination defects. The Libri lab will use Northern blots of reporter genes to assess whether transcription proceeds beyond the PAS sites in cells carrying the *sen1-3* allele. They will also use a method known as CRAC (*in vivo* crosslinking and analysis of cDNA to RNA Pol II) to determine whether the substrate utilization of RNA Pol II is altered (Granneman et al., 2009), indirectly assessing the transcription termination efficiency in *sen1-3* cells. Preliminary result from the Libri lab suggests that *sen1-3*, unlike *sen1-1*, has **no defects** in transcription termination of RNA Pol II genes. Moreover, recent

experiments from our lab has demonstrated that the Sen1-3 protein interacts with RNA Pol II, similar to its wildtype counterpart. Cells carrying the *sen1-3* mutation will also be tested against wildtype and *sen1-1* cells for their ability to spontaneously form R-loops, using antibodies that specific recognize R-loops (S9.6).

Another feature of the *sen1-3* allele that needs addressing is to determine to what extent the corresponding mutant protein affects DNA replication specifically. In fact, the lethality of *sen1-3 rnh1Δ rnh201Δ* cells might not be indicative of obstructed forks. To formally test the effect of the *sen1-3* mutation on the progression of DNA replication, genetic assays will be used to investigate whether head-on and co-directional collisions between replication forks and RNA Pol II transcription complexes trigger increased genomic instability (Prado and Aguilera, 2005). On the other hand, cells carrying the *sen1-3* allele and also deleted for *MRC1* are both temperature-sensitive and hyper-sensitive to hydroxyurea (Section 5.11). This suggests that, in the presence of Sen1-3, replication forks require the Mrc1 protein (and the fork protection complex) during replication under stressful conditions. This can be interpreted as a replication-specific defect in *sen1-3* cells.

5.11 THE *SEN1-3* ALLELE INTERACTS GENETICALLY WITH *MRC1*

The *sen1-1* allele is synthetically lethal with several mutants that provoke fork instability and in the absence of genes involved in homologous recombination (Alzu *et al.* 2012). Consequently, the *sen1-3-TAP* allele was tested for its genetic interaction with some of the mutants that trigger *sen1-1* lethality (Lones & De Piccoli, unpublished). Moreover, synthetic defects in the absence of the Pif1 helicases that are involved in DNA replication were also assessed. Most of the mutants tested did not interact genetically with *sen1-3-TAP*. However, *mrc1*Δ, *tof1*Δ *and rad50*Δ interacted with the *sen1-3-TAP* allele (Fig 5.8). The interaction with Rad50 suggests that R-loops that are not removed by Sen1 can lead to the formation of DSBs whilst interactions with the *tof1*Δ and *mrc1*Δ alleles suggest that Sen1 functions at forks. Because the defects observed in the *sen1-3-TAP mrc1*Δ double mutant at 37ºC and in 50 mM hydroxyurea were more profound than those seen in the *sen1-3-TAP tof1*Δ double mutant in 100 mM HU, it was thought that the defects associated with *sen1-3* led to defects in the S phase checkpoint. Strikingly, however, *sen1-3* presented a more obvious defect in the absence of *CTF18* and *MRC1* than in the absence of *RAD53* (Fig 5.8). As Rad53 is the effector kinase of the S phase checkpoint, these results suggest that the defects observed in the *sen1-3* cells only partially reflects defects in the S phase checkpoint.

The temperature-sensitivity of the *sen1-3-TAP mrc1*Δ double mutant could indicate that, at higher temperatures and in the absence of Mrc1, collisions between replication forks and transcription complexes are increased to such an extent that excessive levels of R-loops are enriched on the genome. Given that the Sen1-3 protein is not enriched locally at forks, the RNase H1 and H2 enzymes would seem unable to remove the corresponding excess of R-loops on their own. In order to investigate the phenotype of the *sen1-3-TAP mrc1*Δ double mutant further, cell cycle progression experiments were conducted (Lones & De Piccoli, unpublished). In brief, control and mutant strains were grown at 24ºC, arrested in G_1 and shifted to 37ºC whilst still arrested. The cells were then released in S phase at 37ºC (Fig 5.9). Upon release from G_1-arrest,

SEN1-TAP cells completed S phase within 40 min, where a major G_2 peak can be seen. *sen1-3-TAP* cells also completed S phase within 40 min, indicating that this mutant does not undergo a prolonged S phase. By contrast, *SEN1-TAP mrc1*Δ took 60 min to complete S phase whilst *sen1-3-TAP mrc1*Δ cells only completed DNA replication 70 min post-release from G_1. In the absence of Rad53, *SEN1-TAP* completed DNA replication in 50 min whilst *sen1-3-TAP* cells did so in 60 min. Interestingly, *sen1-3-TAP mrc1*Δ strains had a significant population of cells stalled in G_2 after S phase completion. Taken together, this suggests that Sen1 associates with replisomes to remove barriers that otherwise trigger continued checkpoint arrest. Given the known functions of Sen1, the barriers in question are probably R-loops or actively transcribing complexes. This reinforces the notion that the transcription process and its intermediates can impede fork progression with dangerous consequences for genome integrity.

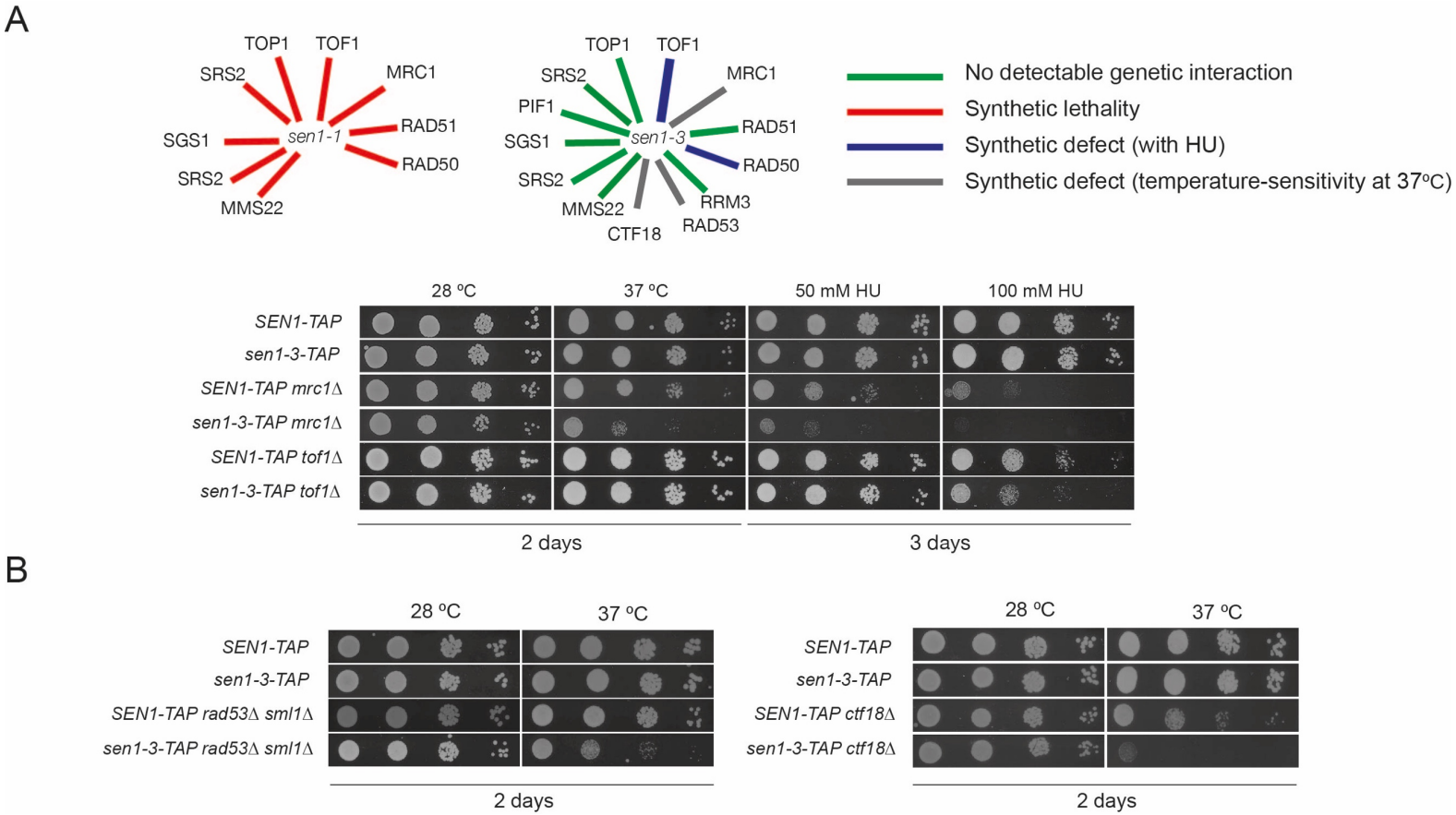


Figure 5.8. Sen1 interacts with *MRC1*. (A) Unlike *sen1-1*, *sen1-3* does not interact genetically with several deletion mutants. However, *sen1-3* interacts profoundly with *mrc1*Δ at 37ºC. **(B)** *sen1-3* displays synthetic defects in the absence of *RAD53* and *CTF18*. However, the synthetic defect with *ctf18*Δ and *mrc1*Δ is stronger than the corresponding defect seen with *rad53*Δ. Lones and De Piccoli (unpublished).

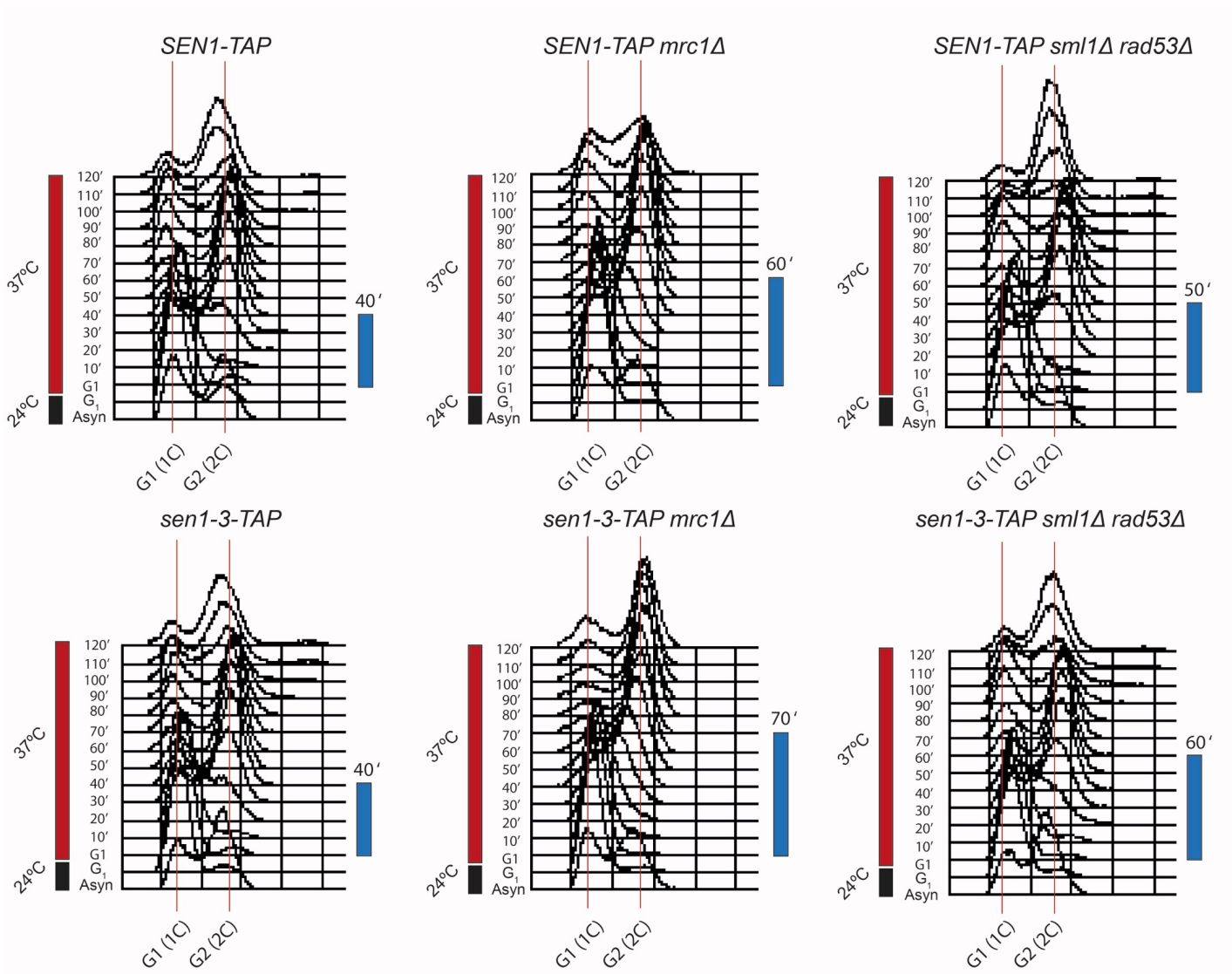

SEN1-TAP
SEN1-TAP mrc1Δ
SEN1-TAP sml1Δ rad53Δ
sen1-3-TAP
sen1-3-TAP mrc1Δ
sen1-3-TAP sml1Δ rad53Δ
37°C
24°C
120′
110′
100′
90′
80′
70′
60′
50′
40′
30′
20′
10′
G1
G1
Asyn
40′
60′
50′
70′
60′
G1 (1C)
G2 (2C)

Figure 5.9. The *sen1-3 mrc1*∆ double mutant has a delayed progression in S phase with a large population of cells arrested in G_2. Cells were grown in YPD at 24ºC and synchronized in G_1. Once arrested, the cells were shifted to 37ºC for 1 h and then released in S phase at 37ºC. FACS samples were collected at the stated time points. The blue bars indicate the duration of replication in minutes. Lones and De Piccoli (unpublished).

5.12 SEN1 USES THE REPLISOME TO REMOVE R-LOOPS AND IS REQUIRED FOR RESPONSE TO FORK BARRIERS

In budding yeast, R-loops are removed by Sen1, RNase H1 and RNase H2 (Aguilera and García-Muse, 2012). Here, we have shown that the triple *rnh1*Δ *rnh201*Δ *sen1-3* mutant is inviable. Double mutants in any combination however are not. Given the overlapping roles of RNase H1 and RNase H2 in the removal of R-loops and the enrichment of these nucleic species upon Sen1 depletion, it is tempting to suggest that Sen1's interaction with replisome components serves to quickly recruit Sen1 at sites where forks encounter R-loops. This localized enrichment of Sen1 facilitates the removal of R-loops.

Overexpression of the *sen1-3* allele using the strong *ACT1* promoter is able to suppress the lethality associated with the triple *rnh1*Δ *rnh201*Δ *sen1-3* mutant. Overexpression of Sen1-3 then must somehow recapitulate (at least partially) the local enrichment of Sen1 at forks. To further characterize the role of Sen1 at forks, a strain encoding the *sen1-3* mutant as well as degron alleles of *RNH1* and *RNH201* is being constructed. This would also help to investigate the role of Sen1 at forks in the absence of RNase H1 and H2 in a single cell cycle. Alternatively, the *rnh1*Δ *rnh201*Δ *ACT1-3HA-sen1-3* that is inviable at 37ºC could be used where cells synchronized in G_1 could be shifted to non-permissive temperatures before release in S phase. It would also be interesting to see whether a *sen1-3* allele with a cloned CIP-tag that recapitulates its interaction with Ctf4 would still lead to synthetic lethality in the absence of both *RNH1* and *RNH201*. The *sen1-3* allele also interacts genetically with both *ctf18*Δ and *mrc*1Δ (Lones and De Piccoli, unpublished). Taken together, these observations strongly suggest that the Sen1 DNA/RNA helicase is required at forks (Fig 5.10).

To further characterize the role of Sen1 at forks, we can exploit the genetic interactions between the *sen1-3* allele and either *ctf18*Δ or *mrc1*Δ, and the fact that overexpression of the RNase H1 gene has been shown to decrease levels of R-loops in several organisms, including yeast (Mischo et al., 2011). Indeed,

should overexpression of RNase H1 suppress the temperature-sensitivity of *sen1-3 ctf18*Δ double mutants and both the temperature-sensitivity and hypersensitivity to HU of *sen1-3 mrc1*Δ double mutants (Fig 5.8), it would strongly suggest that the sole role of Sen1 at forks is R-loop homeostasis. On the other hand, should overexpression of RNase H1 not suppress these phenotypes, it would indicate that Sen1 may have alternative roles at forks.

The fact that Sen1 is required at forks raises an important question on the recruitment process of the protein to forks. It is possible the Sen1 associates with forks at the onset of S phase, irrespective of conditions. Alternatively, it is possible that Sen1 is only recruited when forks encounter R-loops. The latter scenario requires some pathway that would signal Sen1's recruitment at the impeded forks. To differentiate between those two possibilities, immunoprecipitations can be used to monitor the affinity of Sen1 for replisome components in S phase whilst varying the levels of R-loops. For instance, IPs can be carried out in wildtype strains, in strains deleted for both RNase H1 and RNase H2 (favouring stable R-loops) and in strains where RNase H1 is overexpressed (thus quickly removing R-loops). If the affinity of Sen1 for the replisome is independent on the levels of R-loops, it would suggest that Sen1 associates to replisomes, irrespective of whether forks encounter R-loops.

5.13 CONSERVATION OF THE FUNCTION OF SEN1 AT FORKS

In this study, I have shown that Sen1 interacts with replisome components. The utility of this interaction is explored in Sections 5.11 and 5.12. It remains to be seen whether the binding of Sen1 to replisomes and its functioning at forks is conserved. In fungal orthologues of yeast Sen1, of the three residues mutated in *sen1-3*, two tryptophans are invariant (Fig 5.7). This includes the two fission yeast orthologues of Sen1, Dbl8 and Sen1$^{Sz.pombe}$. This suggests that the interaction with replisomes is conserved. In higher eukaryotes, the N-terminal of Senataxin has undergone strong divergence at the level of the primary sequence. However, given that the N-terminal domain of the protein is not separate from its essential, conserved helicase domain, it is possible that mutations that significantly alter the functioning of the protein were selected against. In fact, in human and HeLa cells, Senataxin forms foci upon treatment with aphidicolin, suggesting that Senataxin, like Sen1, also functions at forks (Yüce and West, 2013).

Mutations in the SETX gene causes one of two neurological disorders. One of these disorders, AOA2, can be reconciled with defective transcription termination by Senataxin (Chen et al., 2014). The second disorder, ALS4, has a more obscure pathophysiology. Importantly, ALS4 sufferers are characterized by degeneration of motor neurons but other tissues appear unaffected (Bennett and La Spada, 2015). It is possible that, in human cells, Senataxin travels with forks enabling rapid removal of R-loops, much like Sen1. Moreover, in non-neuronal cells, Senataxin^{ALS4} molecules would appear to compensate for their putative defect in R-loop removal by virtue of still being able to travel with forks, thus providing a localized concentration of the protein where R-loops block fork progression. However, neurons do not replicate their DNA. As such, a similar localized concentration of Senataxin mediated by forks would not be possible in these cells. Consequently, they would accumulate R-loops and DNA damage over time. This would correlate with the progressive nature of the disease. Of the paralogues of Senataxin, mutations in IGHMBP2 has been linked to neuropathy (Guenther et al., 2009) whilst

exogenous RENT1 expression improves survival in nerves harvested from murine models of ALS (Barmada et al., 2015).

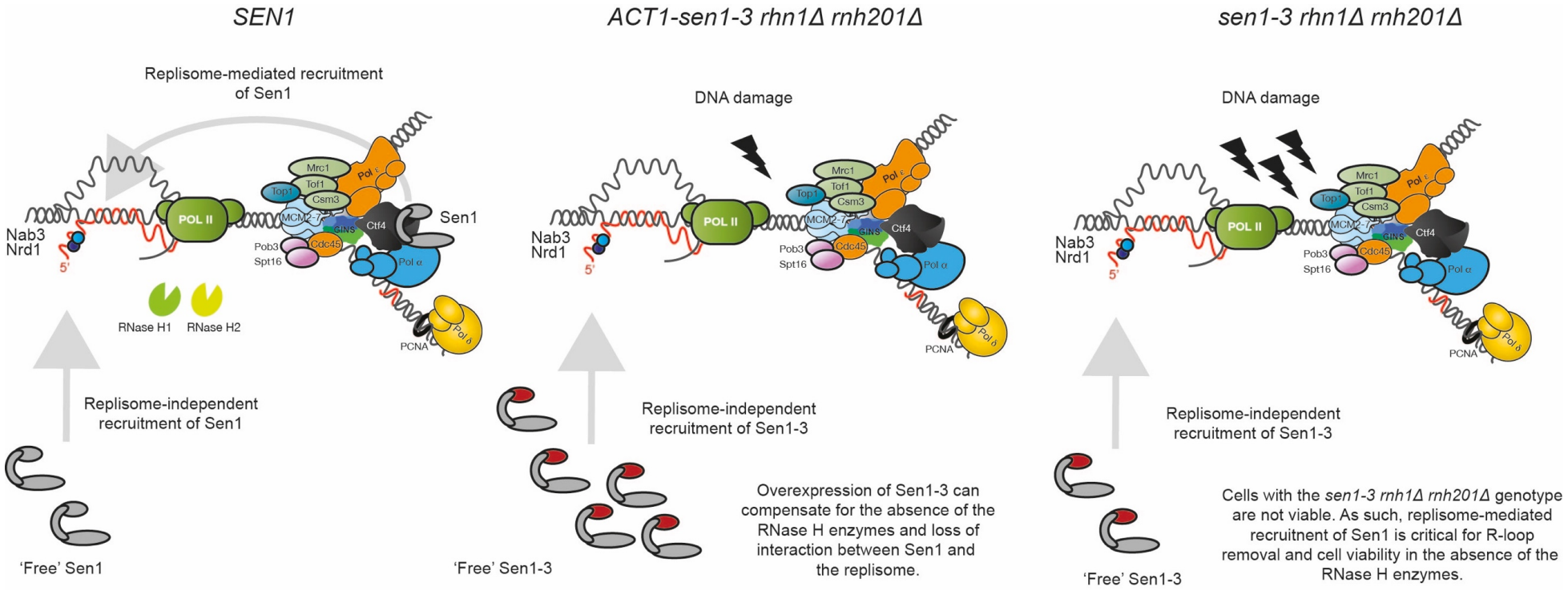


Figure 5.10. The role of the Sen1 DNA/RNA helicase at forks. Results in this study have shed light on the mechanism of action of Sen1 at forks and complement previous studies on the subject (Mischo et al., 2011). In wildtype cells, Sen1 interacts with the replisome by virtue of its N-terminal domain. Sen1 uses this domain to bind to Ctf4 and, perhaps, other replisome components. This enables Sen1 to be quickly recruited at sites of head-on collisions between forks and transcribing complexes, that are favourable to the formation of R-loops (Hamperl et al., 2017). In this study, I have generated a mutant of Sen1, Sen1-3 that does not bind to replisomes. In isolation, this mutant is indistinguishable from wildtype, suggesting that enzymes are able to complement fully the loss of interaction between Sen1 and the replisome. However, in the absence of both RNase H1 and H2, *sen1-3* cells are not viable,

suggesting that the RNase H1 and H2 enzymes can compensate for the loss of interaction between Sen1 and forks. Given the roles of the RNase H enzymes in R-loop removal, this suggests that Sen1 travels with forks to rid the genome of R-loops. Interestingly, overexpression of Sen1-3 using the *ACT1* promoter is able to suppress the lethality of the *sen1-3* allele in the absence of the RNase H enzymes. This suggests that Sen1 can be recruited to sites of collisions between forks and transcribing complexes independently of the replisome.

5.14 CONCLUSION AND PERSPECTIVES

The work in this thesis described the efforts to characterize how eukaryotic cells deal with DNA/RNA hybrids especially in the context of DNA replication. To this end, the role of Pol1 in imprinting in *Sz. pombe* and the role of the DNA/RNA helicase Sen1 at forks in *S. cerevisiae* have been investigated.

For the Pol1 part of the study, a reductive approach was used whereby recombinant proteins were used to differentiate between wildtype Pol1 and its mutant counterpart (Swi7-1) that is defective for imprinting and, thus, cell-type switching in *Sz. pombe* (Gutz and Schmidt, 1985). The *in vitro* results suggests that the mutant protein has reduced affinity for its substrates. Whether this also holds true in an *in vivo* setting remains to be seen but can be determined empirically (Donovan et al., 1997, Liang and Stillman, 1997). An interesting question that arises from these experiments is how do cells cope with a Pol1 variant that has a markedly reduced affinity for its substrates? It is possible that in an *in vivo* context, components of the replisome (such as $Mcl1^{Ctf4}$) serve to alleviate the defect associated with Swi7-1. Alternatively, there might be a mechanism that allows for decoupling of priming by the primase subunit of Pol α and DNA synthesis by its Pol1 subunit where DNA synthesis is instead dependent on Pol δ or ε. In such a situation, the processive polymerases would affect imprinting at the *MPS1* locus. It would be interesting to see whether this potential decoupling also occurs in other eukaryotes.

Meanwhile, in this study we also provide evidence that the DNA/RNA helicase Sen1 travels with replication forks in S phase. We have shown that Sen1 interacts with at least two components of the replisome, including Ctf4 although we have yet to show that Sen1 interacts directly with these proteins. To that end, we need to purify Sen1 and its replisome interactors separately (perhaps from two different host organisms) and reconstitute the putative interaction *in vitro*. Having established that the N-terminal domain of Sen1 is required for interaction with the replisome, we then generated a mutant of Sen1, Sen1-3, that no longer interacts with the replisome. Importantly, the

residues mutated in Sen1-3 are conserved in close homologs of Sen1, including in the Sen1 and Dbl8 proteins from *Sz. pombe*, suggesting that the interaction between Sen1 and the replisome in conserved.

In cells carrying the *sen1-3* allele, absence of both RNase H1 and H2 is lethal, suggesting that Sen1 is required at forks to remove R-loops. This lethality can be suppressed if the *sen1-3* allele is expressed from a strong constitutive (*ACT1*) promoter instead of the endogenous *SEN1* promoter, suggesting that Sen1 can also be recruited in a replisome-independent manner to loci where forks encounter R-loops. There is also synthetic interaction between both *mrc1Δ* and *ctf18Δ* and the *sen1-3* allele, further indicating a role for Sen1 in fork progression. It should be noted that the human homolog of Sen1, Senataxin has been shown to form foci in S phase and upon treatment of the cells with aphidicolin (Yüce and West, 2013). As such, it seems that the role of Sen1 at forks is at least partially conserved in eukaryotes, whereby the replisome actively makes use of the DNA/RNA helicase to remove R-loops. At present, it is also evident that patients with AOA2 and ALS4 present few non-neuronal symptoms. Perhaps this selective penetrance of these diseases reflect the fact that neurons do not undergo DNA replication. It would be interesting to see whether non-neuronal quiescent cells in AOA2 or ALS4 individuals also present defects associated to prolonged presence of R-loops on chromosomes.

SUPPLEMENTARY DATA

Table S.1. Raw data from Mass Spectrometric screen of Sen1 (2-931) IPs (Refer to Sections 4.5- 4.7). Strains used: CS**1852** and CS**1957**.

Prey			**Bait**			
			TAP-Ø	TAP-Sen1 (2-931)		
Gene Name/ Identifier	Accession Number	Molecular Weight	Asynchronous	G1	S	G2
SEN1	YLR430W	253 kDa	0	348	509	340
Retrotransposon TYA Gag and TYB Pol genes	YPL257W-B	199 kDa	13	290	314	248
RPO31	YOR116C	162 kDa	0	282	328	228
RET1	YOR207C	129 kDa	0	183	190	180
CTF4	YPR135W	104 kDa	0	93	243	107
SSB2	YNL209W (+1)	67 kDa	84	76	119	149
TEF2	YBR118W (+1)	50 kDa	86	79	106	106
CDC19	YAL038W	55 kDa	59	57	93	103
SPT16	YGL207W	119 kDa	0	16	233	54
MCM4	YPR019W	105 kDa	0	5	244	45
MCM6	YGL201C	113 kDa	0	0	254	51
RPC82	YPR190C	74 kDa	0	99	90	95
TOF1	YNL273W	141 kDa	0	0	235	42
MRC1	YCL061C	124 kDa	0	8	231	36
MCM5	YLR274W	86 kDa	0	0	192	66
SSA1	YAL005C	70 kDa	63	57	63	82
MCM2	YBL023C	99 kDa	0	0	202	37

POL1	YNL102W	167 kDa	0	2	212	22
MCM3	YEL032W	108 kDa	0	5	184	38
URA2	YJL130C	245 kDa	57	42	55	37
MCM7	YBR202W	95 kDa	0	2	145	29
RPG1	YBR079C	110 kDa	28	30	48	44
ACC1	YNR016C	250 kDa	39	45	38	25
DIA2	YOR080W	85 kDa	0	0	106	34
POL2	YNL262W	256 kDa	0	0	122	10
RPL3	YOR063W	44 kDa	23	20	49	37
PAB1	YER165W	64 kDa	24	26	41	32
GUS1	YGL245W	81 kDa	28	31	28	35
RPS3	YNL178W	27 kDa	27	30	28	30
RPS4B	YHR203C (+1)	29 kDa	20	28	36	33
HSP60	YLR259C	61 kDa	31	27	24	30
SCP160	YJL080C	135 kDa	18	22	45	26
CDC53	YDL132W	94 kDa	0	0	73	33
RPL7A	YGL076C	28 kDa	21	30	27	26
RPC53	YDL150W	47 kDa	0	33	33	35
RPB2	YOR151C	139 kDa	0	28	43	33
CDC45	YLR103C	74 kDa	0	0	80	16
YEF3	YLR249W	116 kDa	23	20	26	26
TDH3	YGR192C	36 kDa	14	24	29	26
RPC40	YPR110C	38 kDa	0	38	27	29
POB3	YML069W	63 kDa	0	2	74	16
RPL4A	YBR031W	39 kDa	19	21	24	29
RPS1B	YML063W	29 kDa	19	21	26	26

DEF1	YKL054C	84 kDa	33	21	24	12
RPL8A	YHL033C	28 kDa	16	20	29	27
RPL26B	YGR034W (+1)	14 kDa	14	20	23	30
RPO21	YDL140C	192 kDa	0	24	38	24
RPC34	YNR003C	36 kDa	0	33	24	29
PRT1	YOR361C	88 kDa	19	19	19	30
SSE1	YPL106C	77 kDa	17	19	14	34
RPS7A	YOR096W	22 kDa	16	20	22	22
Retrotransposon TYA Gag and TYB Pol genes	YCL019W (+5)	202 kDa	0	127	138	109
DED1	YOR204W	66 kDa	19	17	16	25
RPL11B	YGR085C (+1)	20 kDa	21	18	20	15
RPL2A	YFR031C-A (+1)	27 kDa	14	16	20	24
RPS24A	YER074W (+1)	15 kDa	14	15	20	21
RPL17A	YKL180W	21 kDa	12	17	20	19
RPS11B	YBR048W (+1)	18 kDa	18	10	21	22
RPS5	YJR123W	25 kDa	16	17	16	21
RPL9B	YNL067W	22 kDa	19	19	18	13
ASC1	YMR116C	35 kDa	3	21	23	21
RPL20A	YMR242C (+1)	20 kDa	14	15	15	26
CYS4	YGR155W	56 kDa	7	17	20	24
RPL13A	YDL082W (+1)	23 kDa	13	13	20	19
MKT1	YNL085W	94 kDa	13	17	17	19
RPL6A	YML073C	20 kDa	8	19	15	22
UGP1	YKL035W	56 kDa	6	14	18	25
PRI2	YKL045W	62 kDa	0	0	53	5
ATP1	YBL099W	59 kDa	8	13	5	18

RPS6B	YBR181C (+1)	27 kDa	9	12	18	23
RPL10	YLR075W	25 kDa	7	8	18	21
RPS9B	YBR189W (+1)	22 kDa	10	13	24	13
RPL5	YPL131W	34 kDa	12	14	18	17
RPS2	YGL123W	27 kDa	9	10	25	14
TIF3	YPR163C	49 kDa	15	11	17	17
RPS19A	YOL121C	16 kDa	12	12	15	16
RPC17	YJL011C	19 kDa	0	27	13	18
RPC37	YKR025W	32 kDa	0	22	20	16
ACT1	YFL039C	42 kDa	7	15	12	18
SSC1	YJR045C	71 kDa	17	17	12	12
RPP0	YLR340W	34 kDa	8	16	15	14
RPC25	YKL144C	24 kDa	0	23	17	12
RPS13	YDR064W	17 kDa	8	11	16	14
RPL16B	YNL069C	22 kDa	10	10	18	13
CAM1	YPL048W	47 kDa	15	7	12	18
RPS0B	YLR048W	28 kDa	10	13	10	16
RPL21A	YBR191W	18 kDa	6	12	14	16
RPB5	YBR154C	25 kDa	0	21	11	18
SHM2	YLR058C	52 kDa	7	8	16	20
RPL31A	YDL075W	13 kDa	6	8	15	19
ACS2	YLR153C	75 kDa	12	7	11	15
RPL32	YBL092W	15 kDa	7	8	13	15
HTB2	YBL002W	14 kDa	4	3	26	10
ILV2	YMR108W	75 kDa	12	9	9	18
PIL1	YGR086C	38 kDa	6	13	15	16

TFP1	YDL185W	119 kDa	13	9	9	16
POL12	YBL035C	79 kDa	0	0	43	4
GFA1	YKL104C	80 kDa	0	9	19	16
RPS16B	YDL083C (+1)	16 kDa	10	14	13	9
RPL14B	YHL001W (+1)	15 kDa	6	12	13	14
HSC82	YMR186W	81 kDa	11	5	13	16
TOP1	YOL006C	90 kDa	0	0	45	0
RPL33B	YOR234C (+1)	12 kDa	7	8	13	13
OKP1	YGR179C	47 kDa	0	12	16	14
ARO1	YDR127W	175 kDa	0	11	16	12
RPL35B	YDL136W (+1)	14 kDa	5	2	15	15
RPN2	YIL075C	104 kDa	9	7	14	11
ADE3	YGR204W	102 kDa	6	8	17	11
RPS25A	YGR027C (+1)	12 kDa	9	9	9	13
PSF2	YJL072C	25 kDa	0	9	12	21
RPL19B	YBL027W (+1)	22 kDa	7	7	16	11
EFT2	YDR385W (+1)	93 kDa	6	10	11	15
LSP1	YPL004C	38 kDa	9	10	18	16
RPL24B	YGR148C	18 kDa	7	6	12	15
RPS20	YHL015W	14 kDa	6	12	10	11
PSF1	YDR013W	24 kDa	0	9	22	8
PFK2	YMR205C	105 kDa	5	7	18	10
PRI1	YIR008C	48 kDa	0	0	35	5
NIP1	YMR309C	93 kDa	0	12	10	17
RPL36B	YPL249C-A	11 kDa	4	4	11	14
GAL10	YBR019C	78 kDa	7	8	8	12

PDC1	YLR044C	61 kDa	7	8	7	14
KAR2	YJL034W	74 kDa	8	16	14	15
RPL12B	YDR418W (+1)	18 kDa	10	11	9	7
SPT5	YML010W	116 kDa	0	12	14	6
SUP35	YDR172W	77 kDa	9	10	6	10
STM1	YLR150W	30 kDa	11	9	8	8
RPL18B	YNL301C (+1)	21 kDa	3	7	12	12
RPL23A	YBL087C (+1)	14 kDa	11	6	8	9
THS1	YIL078W	85 kDa	7	5	9	14
RPL27A	YHR010W	16 kDa	5	8	12	7
TIF4631	YGR162W	107 kDa	5	7	14	8
TEF4	YKL081W	47 kDa	9	4	9	15
NPL3	YDR432W	45 kDa	7	9	8	8
RPL40A	YIL148W (+1)	15 kDa	3	6	12	3
KGD1	YIL125W	114 kDa	7	6	6	9
RPS27B	YHR021C (+1)	9 kDa	7	4	8	8
MIF2	YKL089W	62 kDa	0	13	8	8
ADH1	YOL086C (+1)	37 kDa	4	10	8	9
CCT8	YJL008C	62 kDa	3	11	4	14
RPS14A	YCR031C	15 kDa	8	8	7	7
RPL25	YOL127W	16 kDa	4	5	8	11
PFK1	YGR240C	108 kDa	5	4	11	14
RPS18A	YDR450W (+1)	17 kDa	3	7	11	8
RPL43B	YJR094W-A (+1)	10 kDa	3	3	9	14
GCD11	YER025W	58 kDa	7	5	4	13
RPN1	YHR027C	109 kDa	0	10	14	6

RPS8A	YBL072C (+1)	22 kDa	0	6	8	14
RPA190	YOR341W	186 kDa	0	4	16	12
DBP2	YNL112W	61 kDa	2	10	5	13
BDP1	YNL039W	68 kDa	0	25	0	3
RPS15	YOL040C	16 kDa	4	7	10	6
SLD5	YDR489W	34 kDa	0	7	13	7
YGR054W	YGR054W	71 kDa	8	4	4	11
RPC31	YNL151C	28 kDa	0	9	7	10
PBP1	YGR178C	79 kDa	14	4	0	7
ILS1	YBL076C	123 kDa	2	5	10	10
NOP58	YOR310C	57 kDa	0	7	2	2
RPS30A	YLR287C-A (+1)	7 kDa	3	4	9	5
SBP1	YHL034C	33 kDa	4	7	6	9
SYP1	YCR030C	96 kDa	0	3	10	11
CSM3	YMR048W	36 kDa	0	0	21	6
FAS1	YKL182W	229 kDa	0	13	7	3
CSE4	YKL049C	27 kDa	0	2	10	13
NOP1	YDL014W	34 kDa	0	5	12	9
SUP45	YBR143C	49 kDa	2	5	7	11
HHT1	YBR010W (+1)	15 kDa	4	0	11	3
BCY1	YIL033C	47 kDa	3	4	9	9
RPA135	YPR010C	136 kDa	2	4	8	10
Retrotransposon TYA Gag and TYB Pol genes	YPR137C-B	199 kDa	0	132	110	67
RPL28	YGL103W	17 kDa	3	0	11	4
RPS10A	YOR293W	13 kDa	3	7	7	4
RPL15A	YLR029C	24 kDa	2	4	4	9

LAT1	YNL071W	52 kDa	3	6	5	8
FKS1	YLR342W	215 kDa	2	6	4	7
RPC19	YNL113W	16 kDa	0	8	4	8
DPB2	YPR175W	78 kDa	0	0	18	5
RPL1B	YGL135W (+1)	24 kDa	0	2	6	6
RPL30	YGL030W	11 kDa	2	5	7	6
RPS23A	YGR118W (+1)	16 kDa	3	2	10	5
GAL1	YBR020W	58 kDa	0	7	4	9
FUN12	YAL035W	112 kDa	2	5	7	6
UTP10	YJL109C	200 kDa	0	0	10	8
SRP54	YPR088C	60 kDa	0	9	4	8
CCT2	YIL142W	57 kDa	3	5	3	8
NEW1	YPL226W	134 kDa	0	6	7	8
ENO1	YGR254W (+1)	47 kDa	2	7	6	5
RPS26B	YER131W (+1)	13 kDa	2	2	4	4
RPS12	YOR369C	15 kDa	0	3	5	3
RPS1A	YLR441C	29 kDa	12	21	26	20
SEC21	YNL287W	105 kDa	3	3	7	6
EFB1	YAL003W	23 kDa	6	0	9	5
GCN1	YGL195W	297 kDa	3	3	3	0
RPL8B	YLL045C	28 kDa	15	0	35	26
RVS161	YCR009C	30 kDa	0	2	4	11
SHM1	YBR263W	54 kDa	5	6	5	8
HRP1	YOL123W	60 kDa	5	7	4	2
TIM44	YIL022W	49 kDa	4	2	2	11
CPR6	YLR216C	42 kDa	0	5	4	9

SSA2	YLL024C	69 kDa	0	48	20	41
HTA2	YBL003C (+1)	14 kDa	0	2	6	5
HHF1	YBR009C (+1)	11 kDa	0	0	10	3
SRV2	YNL138W	58 kDa	6	3	2	4
RRP5	YMR229C	193 kDa	0	2	5	5
RPS17B	YDR447C (+1)	16 kDa	3	4	5	6
Retrotransposon TYA Gag and TYB Pol genes	YMR045C	199 kDa	0	41	30	12
RPB8	YOR224C	17 kDa	0	5	9	4
IMD3	YLR432W	57 kDa	3	6	3	6
LYS21	YDL131W	49 kDa	2	3	3	10
VMA2	YBR127C	58 kDa	0	8	3	6
RPS22A	YJL190C (+1)	15 kDa	2	3	6	5
HSP104	YLL026W	102 kDa	2	2	4	8
NBA1	YOL070C	56 kDa	0	9	4	3
TIF34	YMR146C	39 kDa	0	3	4	8
UTP4	YDR324C	88 kDa	3	0	4	5
ENP1	YBR247C	55 kDa	0	4	3	9
CAJ1	YER048C	45 kDa	4	4	4	4
FAS2	YPL231W	207 kDa	0	7	5	0
Retrotransposon TYA Gag and TYB Pol genes	YMR050C	199 kDa	0	143	0	62
NTH1	YDR001C	86 kDa	3	4	3	5
SEC27	YGL137W	99 kDa	0	2	8	5
FAF1	YIL019W	39 kDa	0	2	0	0
TCP1	YDR212W	60 kDa	0	7	0	7
NOP56	YLR197W	57 kDa	0	4	8	3
COP1	YDL145C	136 kDa	0	0	6	4

RPA49	YNL248C	47 kDa	0	0	4	11
KEM1	YGL173C	175 kDa	2	2	5	4
DHH1	YDL160C	58 kDa	4	2	3	6
YMR315W	YMR315W	38 kDa	0	2	6	6
RNR1	YER070W	100 kDa	3	0	6	4
RPB10	YOR210W	8 kDa	0	0	4	2
FAA1	YOR317W	78 kDa	2	3	2	7
MES1	YGR264C	86 kDa	5	0	4	5
RRM3	YHR031C	82 kDa	0	2	0	0
VMA4	YOR332W	26 kDa	2	0	4	4
RPL34A	YER056C-A (+1)	14 kDa	2	2	7	2
RPL16A	YIL133C	22 kDa	0	5	18	10
GPM1	YKL152C	28 kDa	2	0	2	7
SIS1	YNL007C	38 kDa	0	2	4	6
KES1	YPL145C	49 kDa	2	2	5	5
RPL37A	YLR185W	10 kDa	0	0	6	6
IKI3	YLR384C	153 kDa	0	2	3	6
BFR1	YOR198C	55 kDa	0	5	4	3
HIS4	YCL030C	88 kDa	2	0	6	5
IDP1	YDL066W	48 kDa	0	0	3	10
TAL1	YLR354C	37 kDa	4	0	0	7
ARC1	YGL105W	42 kDa	0	2	5	5
MDH3	YDL078C	37 kDa	3	2	4	4
PDA1	YER178W	46 kDa	0	2	2	7
NAB3	YPL190C	90 kDa	0	4	3	4
RPT1	YKL145W	52 kDa	0	2	3	5

LSG1	YGL099W	73 kDa	2	0	0	8
AME1	YBR211C	37 kDa	0	6	0	6
TYS1	YGR185C	44 kDa	0	0	3	8
BRF1	YGR246C	67 kDa	0	9	0	3
RPS29B	YDL061C	7 kDa	0	2	0	5
RPC10	YHR143W-A	8 kDa	0	5	2	2
RHO1	YPR165W	23 kDa	0	2	4	3
Retrotransposon TYA Gag gene co-transcribed with TYB Pol	YGR109W-A (+4)	34 kDa	0	5	2	3
RPS28A	YOR167C	8 kDa	0	2	0	7
TSL1	YML100W	123 kDa	4	2	4	0
PAT1	YCR077C	88 kDa	5	2	2	0
TRM1	YDR120C	64 kDa	0	3	2	4
GVP36	YIL041W	37 kDa	2	0	2	8
VPS1	YKR001C	79 kDa	4	3	0	5
RPT3	YDR394W	48 kDa	0	3	3	6
CCT4	YDL143W	58 kDa	0	6	0	6
CDC3	YLR314C	60 kDa	2	3	2	3
KRE33	YNL132W	119 kDa	0	2	5	2
SEC26	YDR238C	109 kDa	0	2	4	2
RPL6B	YLR448W	20 kDa	7	0	8	17
CTF19	YPL018W	43 kDa	0	2	3	4
VAS1	YGR094W	126 kDa	0	4	5	2
GLO3	YER122C	55 kDa	0	0	0	3
MCK1	YNL307C	43 kDa	0	0	2	4
SRP1	YNL189W	60 kDa	0	4	2	4
TIF35	YDR429C	31 kDa	0	2	2	5

FBA1	YKL060C	40 kDa	0	0	4	3
ABF2	YMR072W	22 kDa	0	0	3	4
PYC1	YGL062W	130 kDa	0	2	5	0
MSC6	YOR354C	80 kDa	0	3	0	3
RPT6	YGL048C	45 kDa	0	0	6	5
CKA2	YOR061W	39 kDa	0	0	3	7
ATP2	YJR121W	55 kDa	0	6	0	3
QCR2	YPR191W	40 kDa	0	0	2	8
WRS1	YOL097C	49 kDa	0	2	0	8
MDN1	YLR106C	559 kDa	0	0	0	5
PGM1	YKL127W	63 kDa	0	0	0	6
PUF4	YGL014W	98 kDa	0	2	0	4
ARB1	YER036C	68 kDa	2	3	0	3
RPS7B	YNL096C	22 kDa	0	0	0	14
FAA4	YMR246W	77 kDa	0	2	0	6
Retrotransposon TYA Gag and TYB Pol genes	YDR098C-B	199 kDa	0	0	51	66
RVB1	YDR190C	50 kDa	0	0	2	6
GPH1	YPR160W	103 kDa	5	0	0	2
GAL7	YBR018C	42 kDa	0	0	2	6
RPO26	YPR187W	18 kDa	0	5	0	2
SPT15	YER148W	27 kDa	0	3	0	5
CHL4	YDR254W	53 kDa	0	0	0	7
TUB1	YML085C	50 kDa	0	3	0	2
KGD2	YDR148C	50 kDa	2	2	0	3
LSC2	YGR244C	47 kDa	0	2	0	4
PIM1	YBL022C	127 kDa	0	0	4	2

PSF3	YOL146W	22 kDa	0	0	7	0
VIP1	YLR410W	130 kDa	0	3	3	0
RFA1	YAR007C	70 kDa	0	0	6	0
CCT7	YJL111W	60 kDa	0	5	0	0
TUB2	YFL037W	51 kDa	0	3	0	5
ATG9	YDL149W	115 kDa	0	6	0	0
RTN1	YDR233C	33 kDa	2	0	2	2
TSR1	YDL060W	91 kDa	0	0	2	2
CDC12	YHR107C	47 kDa	0	2	0	4
YRA1	YDR381W	25 kDa	0	0	2	3
ARX1	YDR101C	65 kDa	2	0	0	2
RVS167	YDR388W	53 kDa	2	0	0	0
IDH1	YNL037C	39 kDa	2	0	0	3
CDC33	YOL139C	24 kDa	0	0	0	4
ARO80	YDR421W	108 kDa	0	0	3	2
PDB1	YBR221C	40 kDa	0	0	0	4
CCT3	YJL014W	59 kDa	0	3	0	3
DPB3	YBR278W	23 kDa	0	0	3	2
YPK1	YKL126W	76 kDa	0	0	2	3
RPT2	YDL007W	49 kDa	0	0	2	4
RLI1	YDR091C	68 kDa	0	3	0	3
MCM21	YDR318W	43 kDa	0	2	0	4
CLU1	YMR012W	145 kDa	0	0	0	3
PUF6	YDR496C	75 kDa	0	0	2	4
CCT6	YDR188W	60 kDa	0	0	0	2
BNI1	YNL271C	220 kDa	0	0	0	3

SVL3	YPL032C	92 kDa	4	0	0	0
CTF3	YLR381W	84 kDa	0	4	0	0
TPS2	YDR074W	103 kDa	0	0	4	0
HEM15	YOR176W	45 kDa	0	0	0	5
EHT1	YBR177C	51 kDa	0	0	0	5
NET1	YJL076W	129 kDa	0	0	5	0
MEC1	YBR136W	273 kDa	0	0	5	0
SSZ1	YHR064C	58 kDa	2	0	0	2
EGD2	YHR193C	19 kDa	0	0	2	3
ETR1	YBR026C	42 kDa	0	0	2	3
RTT103	YDR289C	46 kDa	0	0	0	4
INP53	YOR109W	125 kDa	0	0	3	2
PWP2	YCR057C	104 kDa	0	2	0	2
NFS1	YCL017C	54 kDa	0	0	2	3
NAN1	YPL126W	101 kDa	0	0	0	3
UTP8	YGR128C	80 kDa	0	0	0	2
HSP42	YDR171W	43 kDa	0	2	0	0
RPL38	YLR325C	9 kDa	0	0	0	3
RPL29	YFR032C-A	7 kDa	0	0	5	0
GAL80	YML051W	48 kDa	0	0	2	3
VPS72	YDR485C	91 kDa	0	0	0	2
YHR020W	YHR020W	77 kDa	3	0	0	3
GLT1	YDL171C	238 kDa	0	0	3	3
SPC105	YGL093W	105 kDa	0	2	0	4
CDC14	YFR028C	62 kDa	0	0	3	0
YLH47	YPR125W	52 kDa	0	0	0	2

UBP3	YER151C	102 kDa	0	2	0	0
YDJ1	YNL064C	45 kDa	0	0	0	5
CBF2	YGR140W	112 kDa	0	0	0	5
SST2	YLR452C	80 kDa	0	0	0	5
RUB1	YDR139C	9 kDa	0	0	5	0
RNR2	YJL026W	46 kDa	0	0	0	5
TDH1	YJL052W	36 kDa	0	0	0	15
KRE2	YDR483W	51 kDa	0	3	0	3
SUI3	YPL237W	32 kDa	2	0	0	2
RFC3	YNL290W	38 kDa	0	0	3	2
YGR207C	YGR207C	29 kDa	3	0	0	2
ISW1	YBR245C	131 kDa	3	0	2	0
CMK2	YOL016C	50 kDa	0	0	2	3
MOT2	YER068W	65 kDa	3	0	0	2
SEC18	YBR080C	84 kDa	0	0	0	3
SIR2	YDL042C	63 kDa	0	0	2	3
NAB2	YGL122C	58 kDa	2	0	0	0
GND1	YHR183W	54 kDa	0	0	0	3
ABD1	YBR236C	50 kDa	0	0	0	3
TAF5	YBR198C	89 kDa	4	0	0	0
MSH6	YDR097C	140 kDa	0	0	3	0
YGR250C	YGR250C	90 kDa	2	0	0	0
NCL1	YBL024W	78 kDa	2	0	0	0
CIT1	YNR001C	53 kDa	0	0	0	3
VPS70	REV_YJR126C	92 kDa	0	0	2	0
TID3	YIL144W	80 kDa	0	0	0	4

POL3	YDL102W	125 kDa	0	0	4	0
NDE1	YMR145C	63 kDa	0	0	0	2
ARO4	YBR249C	40 kDa	0	0	0	2
RSP5	YER125W	92 kDa	2	0	0	0
DIS3	YOL021C	114 kDa	0	2	0	0
TIF5	YPR041W	45 kDa	0	0	0	2
RPT5	YOR117W	48 kDa	0	0	2	2
RFC5	YBR087W	40 kDa	0	0	2	2
CEP3	YMR168C	71 kDa	0	0	0	3
TMA46	YOR091W	40 kDa	0	2	0	0
MIS1	YBR084W	106 kDa	4	0	0	0
TPA1	YER049W	74 kDa	0	0	0	2
NOG1	YPL093W	74 kDa	0	0	0	3
RPL42B	YHR141C (+1)	12 kDa	0	0	0	2
UTP5	YDR398W	72 kDa	0	0	0	3
ZUO1	YGR285C	49 kDa	0	3	0	0
NOG2	YNR053C	55 kDa	0	0	0	2
AMD1	YML035C	93 kDa	0	0	2	0
YSC84	YHR016C	51 kDa	2	0	0	0
DBP8	YHR169W	48 kDa	0	0	0	2
LTE1	REV_YAL024C	163 kDa	0	0	0	3
SMC1	YFL008W	141 kDa	0	0	0	2
RPC11	YDR045C	13 kDa	0	0	0	3
HOC1	YJR075W	46 kDa	0	0	0	2
RPT4	YOR259C	49 kDa	0	4	0	0
KAP123	YER110C	123 kDa	0	0	3	0

CHC1	YGL206C	187 kDa	0	0	0	4
IML3	YBR107C	28 kDa	0	0	0	4
SHS1	YDL225W	63 kDa	0	4	0	0
RAD3	YER171W	90 kDa	0	0	0	4
RVB2	YPL235W	52 kDa	0	0	0	2
PRS1	YKL181W	47 kDa	0	0	0	2
BUD27	YFL023W	91 kDa	0	0	2	0
MGM101	YJR144W	30 kDa	2	0	0	0
ELP3	YPL086C	64 kDa	0	0	0	2
DPB4	YDR121W	22 kDa	0	0	2	0
RPN3	YER021W	60 kDa	0	0	0	2
GLN4	YOR168W	93 kDa	0	0	2	0
POL32	REV_YJR043C	40 kDa	0	0	2	0
PGK1	YCR012W	45 kDa	0	2	0	0
ATP3	YBR039W	34 kDa	0	0	0	2
ALD6	YPL061W	54 kDa	0	2	0	0
ARP2	YDL029W	44 kDa	0	0	0	2
IDH2	YOR136W	40 kDa	0	0	0	2
PRP43	YGL120C	88 kDa	0	0	0	3
NUG1	YER006W	58 kDa	0	0	0	2
ASK10	YGR097W	127 kDa	0	0	2	0
MNN11	YJL183W	48 kDa	0	0	0	2
UTP15	YMR093W	58 kDa	0	0	0	2
MRT4	YKL009W	27 kDa	0	0	0	2
CHS1	YNL192W	130 kDa	0	2	0	0
APL2	YKL135C	82 kDa	0	0	0	2

MAE1	YKL029C	74 kDa	0	0	0	3
GCD10	YNL062C	54 kDa	0	0	0	3
COR1	YBL045C	50 kDa	0	0	0	3
PSA1	YDL055C	40 kDa	0	0	0	3
NAR1	YNL240C	54 kDa	0	0	0	3
CHA1	YCL064C	39 kDa	0	0	0	3
NUF2	YOL069W	53 kDa	0	0	0	3
RFC2	YJR068W	40 kDa	0	0	0	2
YLR225C	YLR225C	46 kDa	0	0	0	2
GCN20	YFR009W	85 kDa	0	0	0	3
RMD9	YGL107C	75 kDa	0	2	0	0
IRC24	YIR036C	29 kDa	0	0	0	2
FAR1	YJL157C	95 kDa	0	2	0	0
SKP1	YDR328C	22 kDa	0	0	2	0
RAD27	YKL113C	43 kDa	0	0	0	2
FRS1	YLR060W	67 kDa	0	0	0	2
ADE12	YNL220W	48 kDa	0	0	0	2
TFB1	YDR311W	73 kDa	2	0	0	0
ACO1	YLR304C	85 kDa	0	0	0	2
MSS51	YLR203C	51 kDa	0	0	0	2
SFI1	YLL003W	113 kDa	2	0	0	0
BRE1	YDL074C	81 kDa	2	0	0	0
MDN1	REV_YLR106C	559 kDa	0	0	0	2
NUP2	YLR335W	78 kDa	2	0	0	0
ILV1	YER086W	64 kDa	0	0	0	2

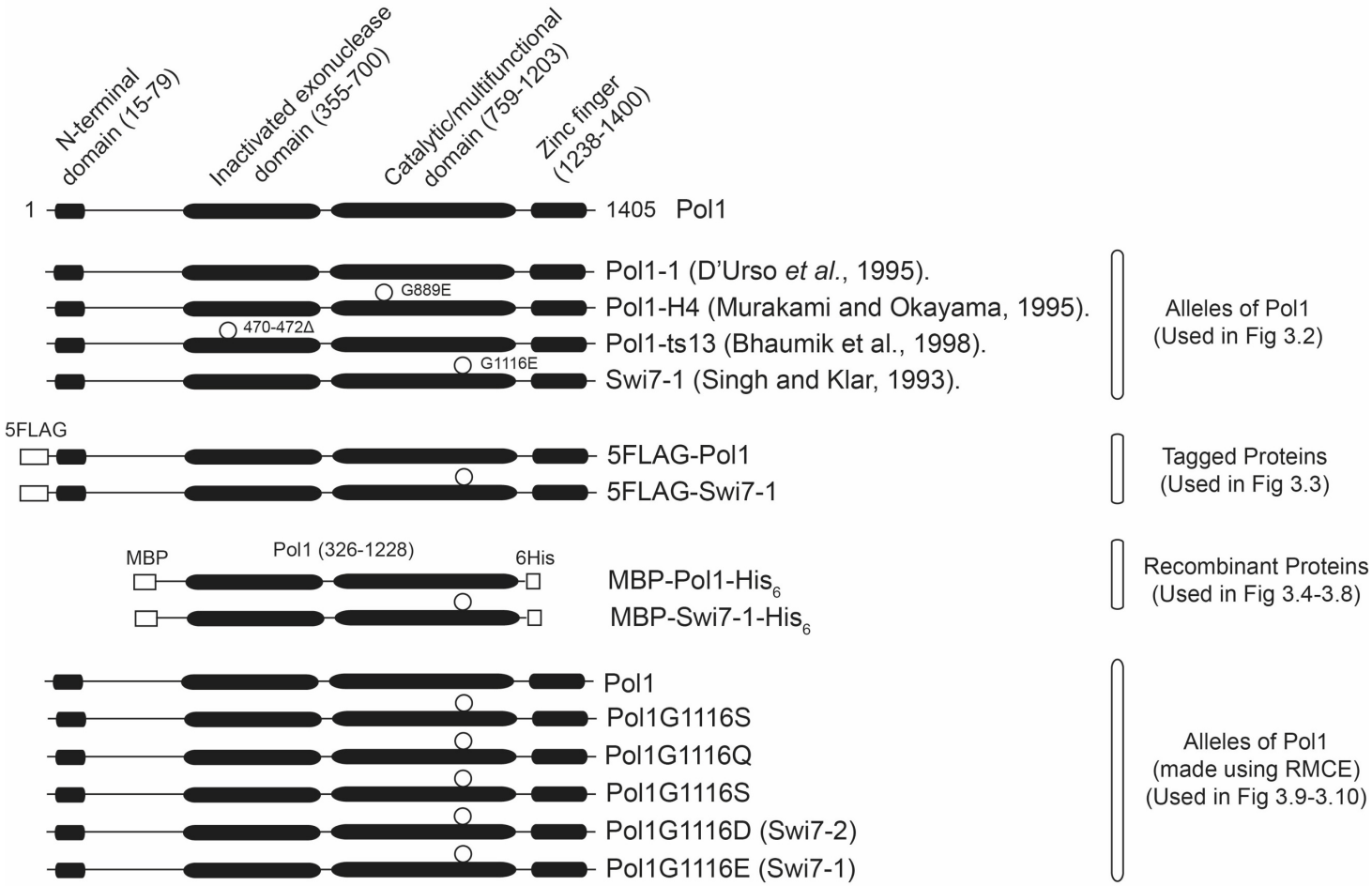
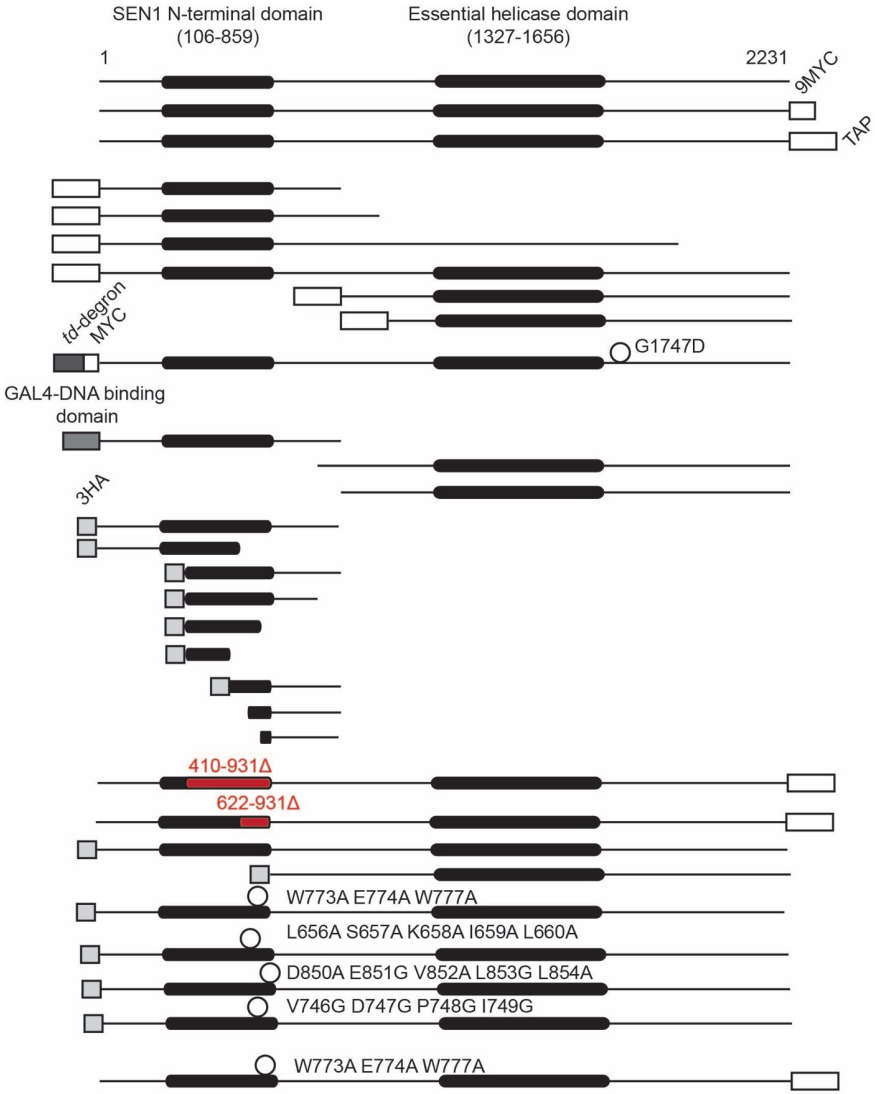


Figure S.1. Schematic of Pol1 mutants and constructs used in this study. Empty circles indicate position of mutations. The mutation in the *pol1-1* allele has not been characterized.



Sen1

Sen1-9MYC (Used in Fig 4.2, 4.18 and 4.19)

Sen1-TAP (Used in Fig 4.3)

TAP-Sen1 (2-931) (Used in Fig 4.6, 4.7, 4.9, 4.10, 4.11, 4.13, 4.15, 4.16, 4.17 and 4.20)

TAP-Sen1 (2-1102) (Used in Fig 4.6 and 4.7)

TAP-Sen1 (2-1901) (Used in Fig 4.6 and 4.7)

TAP-Sen1 (2-2231) (Used in Fig 4.6 and 4.7)

TAP-Sen1 (931-2231) (Used in Fig 4.6 and 4.7)

TAP-Sen1 (1095-2231) (Used in Fig 4.6 and 4.7)

td-MYC-Sen1-1 (Used in Fig 4.7) (Demarini *et al.*, 1992).

GAL4-Sen1 (2-931) (Used in Fig 4.8).

Sen1 (913-2231) (Used in Fig 4.21)

Sen1 (931-2231) (Used in Fig 4.21)

3HA-Sen1 (2-931) (Used in Fig 4.22)

3HA-Sen1 (2-622) (Used in Fig 4.22)

3HA-Sen1 (410-931) (Used in Fig 4.22)

3HA-Sen1 (410-913) (Used in Fig 4.22)

3HA-Sen1 (410-761) (Used in Fig 4.22)

3HA-Sen1 (410-501) (Used in Fig 4.22)

3HA-Sen1 (501-931) (Used in Fig 4.22)

3HA-Sen1 (622-931) (Used in Fig 4.22)

3HA-Sen1 (761-931) (Used in Fig 4.22)

Sen1 (410-931Δ)-TAP (Used in Fig 4.24)

Sen1 (622-931Δ)-TAP (Used in Fig 4.24)

3HA-Sen1 (2-2231) (Used in Fig 4.27, 4.29 and 4.30)

3HA-Sen1 (931-2231) (Used in Fig 4.27 and 4.30)

3HA-Sen1-3 (Used in Fig 4.28, 4.29 and 4.30)

3HA-Sen1-4 (Used in Fig 4.28, 4.29 and 4.30)

3HA-Sen1-5 (Used in Fig 4.28, 4.29 and 4.30)

3HA-Sen1-6 (Used in Fig 4.28, 4.29 and 4.30)

Sen1-3-TAP (Used in Fig 4.31 and 4.32)

Figure S.2. Schematic of Sen1 mutants and constructs used in this study. Empty circles indicate position of point mutations whilst red boxes depict mini-truncations.

ABBREVIATIONS

AGS: Aicardi-Goutières syndrome

ATP: Adenosine triphosphate

APC/C: Anaphase promoting complex/ cyclosome

BLAST: Basic local alignment search tool

BLASTp: Protein-protein BLAST

BSA: Bovine serum albumin

CBP: Calmodulin binding protein

CDK: Cyclin-dependent protein kinase

CF: Cleavage factor

CFS: Common fragile site

ChIP: Chromatin immuno-precipitation

CIP: Ctf4 interacting peptide

CMG: Cdc45 MCM2-7 GINS

CMGE: Cdc45 MCM2-7 GINS Polymerase ε

CPA: Cleavage and polyadenylation

CPF: Cleavage and polyadenylation factor

CPSF: Cleavage and polyadenylation specificity factor

CTD: Carboxyl-terminal domain

CUT: Cryptic unstable transcript

DDK: Dbf4-dependent kinase

DSB: Double-stranded break

dsDNA: Double-stranded DNA

FLC: *FLOWERING LOCUS C*

GAL: Galactose

GINS: Sld5 (Go), Psf1 (Ichi), Psf2 (Ni), Psf3 (San)

HU: Hydroxyurea

KSHV: Kaposi's sarcoma-associated herpesvirus

MAT: Mating-type

MBP: Maltose binding protein

MCM: Minichromosome maintenance

MCS: Multiple cloning site

MPS1: *mat1 pausing site 1*

MMS: Methyl methanesulfonate

mtDNA: Mitochondrial DNA

NER: Nucleotide excision repair

PAS: Polyadenylation signal

PIP: PCNA interacting peptide

PMSF: Phenylmethane sulfonyl fluoride

PSI-BLAST: Position-specific iterative BLAST

PSSM: Position specific score matrix

Raff: Raffinose

RFB: Replication fork barrier

RFC: Replication factor C

RPA: Replication protein A

RPC: Replication progression complex

RTS1: Replication termination site 1

SDSA: Synthesis-dependent strand annealing

ssDNA: Single-stranded DNA

TAM: Transcription-associated mutagenesis

TAR: Transcription-associated recombination

TCR: Transcription coupled repair

TD (or *td*): Temperature degron

TE: Transposable elements

VLP: Virus-like particle

Y2H: Yeast-2-hybrid

REFERENCES

ABID ALI, F., RENAULT, L., GANNON, J., GAHLON, H., L., KOTECHA, A., ZHOU, J., C., RUEDA, D. & COSTA, A. 2016. Cryo-EM structures of the eukaryotic replicative helicase bound to a translocation substrate. *Nature Communications,* 7**,** 10708

AGUILERA, A. & GARCÍA-MUSE, T. 2012. R-Loops: From transcription byproducts to threats to genome stability. *Molecular Cell,* 46**,** 115- 124.

AHMED, S., SAINI, S., ARORA, S. & SINGH, J. 2001. Chromodomain protein Swi6-mediated role of DNA Polymerase α in establishment of silencing in fission yeast. *Journal of Biological Chemistry,* 276**,** 47814-47821.

AHN, A., H., KIM, M. & BURATOWSKI, S. 2004. Phosphorylation of serine 2 within the RNA Polymerase II C-terminal domain couples transcription and 3′ end processing. *Molecular Cell,* 13**,** 67- 76.

AKAMATSU, Y. & KOBAYASHI, T. 2015. The human RNA Polymerase I transcription terminator complex acts as a replication fork barrier that coordinates the progress of replication with rRNA transcription activity. *Molecular and Cellular Biology,* 35**,** 1871- 1881.

ALCASABAS, A., A., OSBORN, A., J., BACHANT, J., HU, F., WERLER, P., J., BOUSSET, K., FURUYA, K., DIFFLEY, J., F., CARR, A., M. & ELLEDGE, S., J. 2001. Mrc1 transduces signals of DNA replication stress to activate Rad53. *Nature Cell Biology,* 3**,** 958- 965

ALLEN, J., B., ZHOU, Z., SIEDE, W., FRIEDBERG, E., C. & ELLEDGE, S., J. 1994. The *SAD1/RAD53* protein kinase controls multiple checkpoints and DNA damage-induced transcription in yeast. *Genes and Development,* 8**,** 2401- 2415.

ALTSCHUL, S., F., MADDEN, T., L., SCHÄFFER, A., A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D., J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* 25**,** 3389-3402.

ALZU, A., BERMEJO, R., BEGNIS, M., LUCCA, C., PICCINI, D., CAROTENUTO, W., SAPONARO, M., BRAMBATI, A., COCITO, A., FOIANI, M. & LIBERI, G. 2012. Senataxin associates with replication forks to protect fork integrity across RNA-Polymerase-II-transcribed genes. *Cell,* 151**,** 835- 846.

ARCANGIOLI, B. 1998. A site- and strand-specific DNA break confers asymmetric switching potential in fission yeast. *The EMBO Journal,* 17**,** 4503- 4510.

ARCANGIOLI, B. & DE LAHONDÈS, R. 2000. Fission yeast switches mating type by a replication-recombination coupled process. *The EMBO Journal,* 19**,** 1389- 1396.

ARCANGIOLI, B. & KLAR, A., J,S. 1991. A novel switch-activating site (S*AS1)* and its cognate binding factor (Sap1) required for efficient m*at1* switching in S*chizosaccharomyces pombe*. *The EMBO Journal,* 10**,** 3025- 3032.

ARIGO, J., T., EYLER, D., E., CARROLL, K., L. & CORDEN, J., L. 2006. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Molecular Cell,* 23**,** 841- 851.

BABA, T., ARA, T., HASEGAWA, M., TAKAI, Y., OKUMURA, Y., BABA, M., DATSENKO, K., A., TOMITA, M., WANNER, B., L. & MORI, H. 2006. Construction of E*scherichia coli* K-12 in-frame single-gene knockout mutants: the Keio collection. *Molecular Systems Biology.***,** 2006.008.

BALDACCI, G., CHERIF-ZAHAR, B. & BERNARDI, G. 1984. The initiation of DNA replication in the mitochondrial genome of yeast. *The EMBO Journal,* 3**,** 2115- 2120.

BALK, B., MAICHER, A., DEES, M., KLERMUND, J., LUKE-GLASER, S., BENDER, K. & LUKE, B. 2013. Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nature Structural Molecular Biology,* 20**,** 1199- 1205.

BANDO, M., KATOU, Y., KOMATA, M., TANAKA, H., ITOH, T., SUTANI, T. & SHIRAHIGE, K. 2009. Csm3, Tof1, and Mrc1 form a heterotrimeric mediator complex that associates with DNA replication forks. *The Journal of Biological Chemistry,* 284**,** 34355- 34365

BARAN, N., LAPIDOT, A. & MANOR, H. 1991. Formation of DNA triplexes accounts for arrests of DNA synthesis at d(TC)n and d(GA)n tracts. *Microbiology and Molecular Biology Reviews,* 71**,** 13– 35.

BARANOVSKIY, A. G., BABAYEVA, N., D ., SUWA, Y., GU, J., PAVLOV, Y., I . & TAHIROV, T. H. 2014. Structural basis for inhibition of DNA replication by aphidicolin. *Nucleic Acids Research,* 42**,** 14013- 14021.

BARLOW, J., H., FARYABI, R., B., CALLÉN, E., WONG, N., MALHOWSKI, A., CHEN, H., T., GUTIERREZ-CRUZ, G., SUN, H., W., MCKINNON, P., WRIGHT, G., CASELLAS, R., ROBBIANI, D., F., STAUDT, L., FERNANDEZ-CAPETILLO, O. & NUSSENZWEIG, A. 2013. Identification of early replicating fragile sites that contribute to genome instability. *Cell,* 152**,** 620- 632.

BARMADA, S., J., JU, S., ARJUN, A., BATARSE, A., ARCHBOLD, H., C., PEISACH, D., LI, X., ZHANG, Y., TANK, E., M., QIU, H., HUANG, E., J., RINGE, D., PETSKO, G., A. & FINKBEINER, S. 2015. Amelioration of toxicity in neuronal models of amyotrophic lateral sclerosis by hUPF1. *Proceedings for the National Academy of Sciences of the USA,* 112**,** 7821- 7826

BARNUM, K., J. & O'CONNELL, M., J. 2014. Cell cycle regulation by checkpoints. *Methods in Molecular Biology,* 1170**,** 29- 40.

BASTIA, D. & ZAMAN, S. 2014. Mechanism and physiological significance of programmed replication termination. *Seminars in Cell and Developmental Biology,* 30**,** 165- 173

BECHEREL, O., J., SUN, J., YEO, A., J., NAYLER, S., FOGEL, B., L., GAO, F., COPPOLA, G., CRISCUOLO, C., DE MICHELE, G., WOLVETANG, E. & LAVIN, M., F. 2015. A new model to study neurodegeneration in ataxia oculomotor apraxia type 2. *Human Molecular Genetcis,* 24**,** 5759–5774

BECHEREL, O., J., YEO, A., J., STELLATI, A., HENG, E., Y., LUFF, J., SURAWEERA, A., M., WOODS, R., FLEMING, J., CARRIE, D., MCKINNEY, K., XU, X., DENG, C. & LAVIN, M., F . 2013. Senataxin

plays an essential role with DNA damage response proteins in meiotic recombination and gene silencing. *PLoS Genetics,* 9**,** e1003435.

BECKER, R., LOLL, B. & MEINHART, A. 2008. Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II. *Journal of Biological Chemistry,* 283**,** 22659- 22669.

BEKAERT, S., DE MEYER, T. & VAN OOSTVELDT, P. 2005. Telomere attrition as ageing biomarker. *Anticancer Research,* 25**,** 3011- 3021.

BELETSKII, A. & BHAGWAT, A., S. 1996. Transcription-induced mutations: Increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. *Proceedings for the National Academy of Sciences of the USA,* 93**,** 13919- 13924.

BENNETT, C., L., CHEN, Y., VIGNALI, M., LO, R., S., MASON, A., G., UNAL, A., HUQ SAIFEE, N., P., FIELDS, S. & LA SPADA, A., R. 2013. Protein interaction analysis of Senataxin and the ALS4 L389S mutant yields insights into Senataxin post-translational modification and uncovers mutant-specific binding with a brain cytoplasmic RNA-encoded peptide. *PLoS ONE,* 8**,** e78837.

BENNETT, C., L. & LA SPADA, A. R. 2015. Unwinding the role of senataxin in neurodegeneration. *Discovery Medicine,* 19**,** 127- 136.

BERMUDEZ, V., P., FARINA, A., TAPPIN, I. & HURWITZ, J. 2010. Influence of the human cohesion establishment factor Ctf4/AND-1 on DNA replication. *The Journal of Biological Chemistry,* 285**,** 9493- 9505.

BHAUMIK, D. & WANG, T., S,F. 1998. Mutational effect of fission yeast Pol α on cell cycle events. *Molecular Biology of the Cell,* 9**,** 2107- 2123.

BIERNE, H., EHRLICH, S., D. & MICHEL, B. 1991. The replication termination signal *terB* of the E*scherichia coli* chromosome is a deletion hot spot. *The EMBO Journal,* 10**,** 2699- 2705.

BLATTNER, F., R., PLUNKETT, G., 3RD., BLOCH, C., A., PERNA, N., T., BURLAND, V., RILEY, M., COLLADO-VIDES, J., GLASNER, J., D., RODE, C., K., MAYHEW, G., F., GREGOR, J., DAVIS, N., W., KIRKPATRICK, H., A., GOEDEN, M., A., ROSE, D., J., MAU, B. & SHAO, Y. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science,* 277**,** 1453- 1462.

BOCHMAN, M., L. & SCHWACHA, A. 2008. The Mcm2-7 complex has *in vitro* helicase activity. *Molecular Cell,* 31**,** 287- 293

BODENMILLER, B., WANKA, S., KRAFT, C., URBAN, J., CAMPBELL, D., PEDRIOLI, P., G., GERRITS, B., PICOTTI, P., LAM, H., VITEK, O., BRUSNIAK, M., Y., ROSCHITZKI, B., ZHANG, C., SHOKAT, K., M., SCHLAPBACH, R., COLMAN-LERNER, A., NOLAN, G., P., NESVIZHSKII, A., I., PETER, M., LOEWITH, R., VON MERING, C. & AEBERSOLD, R. 2010. Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Science Signalling,* 3**,** rs4.

BONNEAU, F., BASQUIN, J., EBERT, J., LORENTZEN, E. & CONTI, E. 2009. The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. *Cell,* 139**,** 547- 559.

BREWER, B., J. 1988. When polymerases collide: replication and transcriptional organization of the *E. coli c*hromosome. *Cell,* 53**,** 679- 686.

BREWER, B., J. & FANGMAN, W., L. 1988. A replication fork barrier at the 3' end of yeast ribosomal RNA genes. *Cell,* 55**,** 637- 643.
BRUCK, I. & KAPLAN, D., L. 2013. Cdc45 protein-single-stranded DNA interaction is important for stalling the helicase during replication stress. *The Journal of Biological Chemistry,* 288**,** 7550– 7563.
BUENO, A. & RUSSELL, P. 1993. Two fission yeast B-type cyclins, Cig2 and Cdc13, have different functions in mitosis. *Molecular and Cellular Biology,* 13**,** 2286- 2297
BURATOWSKI, S. 2003. The CTD code. *Nature Structural Biology,* 10**,** 679-680.
CALZADA, A., HODGSON, B., KANEMAKI, M., BUENO, A. & LABIB, K. 2005. Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork. *Genes and Development,* 19**,** 1905-1919
CARROLL, K., L., GHIRLANDO, R., AMES, J., M. & CORDEN, J., L. 2007. Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA,* 13**,** 361- 373.
CASTEL, A., L., CLEARY, J., D. & PEARSON, C., E. 2010. Repeat instability as the basis for human diseases and as a potential target for therapy. *Nature Reviews Molecular and Cellular Biology,* 11**,** 165- 170.
CASTELLANO-POZO, M., SANTOS-PEREIRA, J., M., RONDÓN, A., G., BARROSO, S., ANDÚJAR, E., PÉREZ-ALEGRE, M., GARCÍA-MUSE, T. & AGUILERA, A. 2013. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Molecular Cell,* 52**,** 583-590
CERRITELLI, S., M. & CROUCH, R., J. 2009. Ribonuclease H: the enzymes in eukaryotes. *The FEBS Journal,* 276**,** 1494- 1505
CHANFREAU, G., ELELA, S., A., ARES, M., JR. & GUTHRIE, C. 1997. Alternative 3'-end processing of U5 snRNA by RNase III. *Genes and Development,* 11**,** 2741- 2751.
CHAUDHURI, J. & ALT, F., W. 2004. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nature Reviews Immunology,* 4**,** 541- 552.
CHAVEZ, S. & AGUILERA, A. 1997. The yeast *HPR1* gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. *Genes and Development,* 11**,** 3459- 3470.
CHEN, X., MÜLLER, U., SUNDLING, K., E. & BROW, D., A. 2014. *Saccharomyces cerevisiae* Sen1 as a model for the study of mutations in human Senataxin that elicit cerebellar ataxia. *Genetics,* 198**,** 577–590.
CHEN, X., POOREY, K., CARVER, M., N., MÜLLER, U., BEKIRANOV, S., AUBLE, D., T. & BROW, D., A. 2017. Transcriptomes of six mutants in the Sen1 pathway reveal combinatorial control of transcription termination across the *Saccharomyces cerevisiae* genome. *PLoS Genetics,* 13**,** e1006863.
CHEN, Y., Z., BENNETT, C., L., HUYNH, H., M., BLAIR, I., P., PULS, I., IROBI, J., DIERICK, I., ABEL, A., KENNERSON, M., L., RABIN, B., A., NICHOLSON, G., A., AUER-GRUMBACH, M., WAGNER, K., DE JONGHE, P., GRIFFIN, J., W., FISCHBECK, K., H., TIMMERMAN, V., CORNBLATH, D., R. & CHANCE, P., F. 2004. DNA/RNA helicase gene

mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *American Journal of Human Genetics,* 74**,** 1128- 1135.

CHILKOVA, O., STENLUND, P., ISOZ, I., STITH, C., M., GRABOWSKI, P., LUNDSTRÖM, E., B., BURGERS, P., M. & JOHANSSON, E. 2007. The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Research,* 35**,** 6588- 6597

CHINCHILLA, K., RODRIGUEZ-MOLINA, J., B., URSIC, D., FINKEL, J., S., ANSARI, A., Z. & CULBERTSON, M., R. 2012. Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in *Saccharomyces cerevisiae*. *Eukaryotic Cell,* 11**,** 417- 429.

CHLEBOWSKI, A., LUBAS, M., JENSEN, T., H. & DZIEMBOWSKI, A. 2013. RNA decay machines: The exosome. *Biochimica et Biophysica Acta,* 1829**,** 552- 560.

CODLIN, S. & DALGAARD, J., Z. 2003. Complex mechanism of site-specific DNA replication termination in fission yeast. *The EMBO Journal,* 22**,** 3431- 3440.

CONNOLLY, T. & BEACH, D. 1994. Interaction between the Cig1 and Cig2 B-type cyclins in the fission yeast cell cycle. *Molecular and Cellular Biology,* 14**,** 768- 776

CONRAD, N., K., WILSON, S., M., STEINMETZ, E., J., PATTURAJAN, M., BROW, D., A., SWANSON, M., S. & CORDEN, J., L. 2000. A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics,* 154**,** 557- 571.

CONTI, C., SACCÀ, B., HERRICK, J., LALOU, C., POMMIER, Y. & BENSIMON, A. 2007. Replication fork velocities at adjacent replication origins are coordinately modified during DNA replication in human cells. *Molecular Biology of the Cell,* 18**,** 3059– 3067.

CORDEN, J., L. 2013. The RNA Polymerase II C-terminal domain: Tethering transcription to transcript and template. *Chemical Reviews,* 113**,** 8423-8455.

COSKUN-ARI, F., F. & HILL, T., M. 1997. Sequence-specific interactions in the Tus-Ter complex and the effect of base pair substitutions on arrest of DNA replication in E*scherichia coli*. *The Journal of Biological Chemistry,* 272**,** 26448- 26456.

CROW, Y., J., VANDERVER, A., ORCESI, S., KUIJPERS , T., W. & RICE , G., I. 2014. Therapies in Aicardi–Goutières syndrome. *Clinical and Experimental Immunology,* 175**,** 1–8

CSORBA, T., QUESTA, J., I., SUN, Q. & DEAN, C. 2014. Antisense *COOLAIR* mediates the coordinated switching of chromatin states at *FLC* during vernalization. *Proceedings for the National Academy of Sciences of the USA,* 111**,** 16160- 16165.

DAIGAKU, Y., KESZTHELYI, A., MÜLLER, C., A ., MIYABE, I., BROOKS, T., RETKUTE, R., HUBANK, M., NIEDUSZYSKI, C., A . & CARR, A., M . 2015. A global profile of replicative polymerase usage. *Nature Structural and Molecular Biology,* 22**,** 192-198.

DALGAARD, J., Z., GODFREY, E., L. & MACFARLANE, R., J. 2011. Eukaryotic replication barriers: How, why and where forks stall. *In:*

SELIGMANN, H. (ed.) *DNA Replication-Current Advances.* Online: InTech.
DALGAARD, J., Z. & KLAR, A., J,S. 1999. Orientation of DNA replication establishes mating-type switching pattern in *S. pombe*. *Nature,* 400**,** 181- 184
DALGAARD, J., Z. & KLAR, A., J,S. 2000. *swi1* and *swi3* perform imprinting, pausing, and termination of DNA replication in *S. pombe*. *Cell,* 102**,** 745- 751.
DALGAARD, J., Z. & KLAR, A., J,S. 2001. A DNA replication-arrest site *RTS1* regulates imprinting by determining the direction of replication at *mat1* in *S. pombe*. *Genes and Development,* 15**,** 2060- 2068.
DANIELS, G., A. & LIEBER, M., R. 1995. RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. *Nucleic Acids Research,* 23**,** 5006- 5011.
DATTA, A. & JINKS-ROBERTSON, S. 1995. Association of increased mutation rates with high levels of transcription in yeast. *Science,* 268**,** 1616- 1619.
DE MARCH, M., MERINO, N., BARRERA-VILARMAU, S., CREHUET, R., ONESTI, S., BLANCO, F., J. & DE BIASIO, A. 2017. Structural basis of human PCNA sliding on DNA. *Nature Communications,* 8.
DELLINO, G., I., CITTARO, D., PICCIONI, R., LUZI, L., BANFI, S., SEGALLA, S., CESARONI, M., MENDOZA-MALDONADO, R., GIACCA, M. & PELICCI, P., G. 2013. Genome-wide mapping of human DNA-replication origins: Levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome Research,* 23**,** 1- 11.
DEMARINI, D., J., PAPA, F., R., SWAMINATHAN, S., URSIC, D., RASMUSSEN, T., P., CULBERTSON, M., R. & HOCHSTRASSER, M. 1995. The yeast *SEN3* gene encodes a regulatory subunit of the 26S proteasome complex required for ubiquitin-dependent protein degradation *in vivo*. *Molecular and Cellular Biology,* 15**,** 6311- 6321.
DEMARINI, D., J., WINEY, M., URSIC, D., WEBB, F. & CULBERTSON, M., R. 1992. *SEN1*, a positive effector of tRNA-splicing endonuclease in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology,* 12**,** 2154-2164.
DESANY, B., A., ALCASABAS, A., A., BACHANT, J., B. & ELLEDGE, S., J. 1998. Recovery from DNA replicational stress is the essential function of the S-phase checkpoint pathway. *Genes and Development,* 12**,** 2956- 2970
DEVBHANDARI, S., JIANG, J., KUMAR, C., WHITEHOUSE, I. & REMUS, D. 2017. Chromatin constrains the initiation and elongation of DNA replication. *Molecular Cell,* 65**,** 131-141.
DING, B., LEJEUNE, D. & LI, S. 2010. The C-terminal repeat domain of Spt5 plays an important role in suppression of Rad26-independent transcription coupled repair. *The Journal of Biological Chemistry,* 285**,** 5317- 5326.
DONOVAN, S., HARWOOD, J., DRURY, L., S. & DIFFLEY, J., F,X. 1997. Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. *Proceedings for the National Academy of Sciences of the USA,* 94**,** 5611– 5616.

DROZDETSKIY, A., COLE, C., PROCTER, J. & BARTON, G., J. 2015. JPred4: a protein secondary structure prediction server. *Nucleic Acids Research,* 43**,** W389- W394

DUA, R., LEVY, D., L. & CAMPBELL, J., L. 1999. Analysis of the essential functions of the C-terminal protein/protein interaction domain of *Saccharomyces cerevisiae* Pol ε and its unexpected ability to support growth in the absence of the DNA polymerase domain. *Journal of Biological Chemistry,* 274**,** 222283- 222288.

DUA, R., LEVY, D., L., LI, C., M., SNOW, P., M. & CAMPBELL, J., L. 2002. I*n vivo r*econstitution of S*accharomyces cerevisiae* DNA polymerase epsilon in insect cells. Purification and characterization. *The Journal of Biological Chemistry,* 277**,** 7889- 7896

D'URSO, G., GRALLERT, B. & NURSE, P. 1995. DNA polymerase alpha, a component of the replication initiation complex is essential for the checkpoint coupling S phase to mitosis in fission yeast. *Journal of Cell Science,* 108**,** 3109- 3118.

EGEL, R. 1989. *Mating-type genes, meiosis, and sporulation,* New York, Harcourt Brace Jovanovich, Publishers.

EL HAGE, A., FRENCH, S., L., BEYER, A., L. & TOLLERVEY, D. 2010. Loss of topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes and Development,* 24**,** 1546- 1558.

EL HAGE, A., KOPER, M., KUFEL, J. & TOLLERVEY, D. 2008. Efficient termination of transcription by RNA polymerase I requires the 5' exonuclease Rat1 in yeast. *Genes and Development,* 22**,** 1069- 1081.

EL HAGE, A., WEBB, S., KERR, A. & TOLLERVEY, D. 2014. Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genetics,* 10**,** e1004716.

ELELA, S., A. & ARES, M., JR. 1998. Depletion of yeast RNase III blocks correct U2 3' end formation and results in polyadenylated but functional U2 snRNA. *The EMBO Journal,* 17**,** 3738- 3746.

ELVERS, I., JOHANSSON, F., GROTH, P., ERIXON, K. & HELLEDAY, T. 2011. UV stalled replication forks restart by re-priming in human fibroblasts. *Nucleic Acids Research,* 39**,** 7049- 7057.

EPSTEIN, C., B. & CROSS, F., R. 1992. *CLB5*: a novel B cyclin found in budding yeast with a role in S phase. *Genes and Development,* 6**,** 1695- 1706.

ERRINGTON, J. 2003. Regulation of endospore formation in *Bacillus subtilis*. *Nature Reviews in Microbiology,* 1**,** 117- 126.

EVANS, T., ROSENTHAL, E., T., YOUNGBLOM, J., DISTEL, D. & HUNT, T. 1983. Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. *Cell,* 33**,** 389- 396.

EVRIN, C., CLARKE, P., ZECH, J., LURZ, R., SUN, J., UHLE, S., LI, H., STILLMAN, B. & SPECK, C. 2009. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proceedings for the National Academy of Sciences of the USA,* 106**,** 20240-20245.

EYDMANN, T., SOMMARIVA, E., INAGAWA, T., MIAN, S., KLAR, A., J,S. & DALGAARD, J., Z. 2008. Rtf1-mediated eukaryotic site-specific replication termination. *Genetics,* 180**,** 27–39.

FALLET, E., JOLIVET, P., SOUDET, J., LISBY, M., GILSON, E. & TEIXEIRA, M., T. 2014. Length-dependent processing of telomeres in the absence of telomerase. *Nucleic Acids Research,* 42**,** 3648- 3665.

FASKEN, M., B., LARIBEE, R., N. & CORBETT, A., H. 2015. Nab3 facilitates the function of the TRAMP complex in RNA processing via recruitment of Rrp6 independent of Nrd1. *PLoS Genetics,* 11**,** e1005044.

FENG, W. & D'URSO, G. 2001. *Schizosaccharomyces pombe* cells lacking the amino-terminal catalytic domains of DNA polymerase ε are viable but require the DNA damage checkpoint control. *Molecular and Cellular Biology,* 21**,** 4495-4504.

FISHER, D., L. & NURSE, P. 1996. A single fission yeast mitotic cyclin B p34cdc2 kinase promotes both S-phase and mitosis in the absence of G1 cyclins. *The EMBO Journal,* 15**,** 850-860.

FORSBURG, S., L. 2004. Eukaryotic MCM proteins: Beyond replication initiation. *Microbiology and Molecular Biology Reviews,* 68**,** 109–131.

FORSBURG, S., L. & RHIND, N. 2006. Basic methods for fission yeast. *Yeast,* 23**,** 173–183.

FOURY, F., ROGANTI, T., LECRENIER, N. & PURNELLE, B. 1998. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Letters,* 440**,** 325-331.

FOX, M., J., GAO, H., SMITH-KINNAMAN, W., R., LIU, Y. & MOSLEY, A., L. 2015. The exosome component Rrp6 is required for RNA polymerase II termination at specific targets of the Nrd1-Nab3 pathway. *PLoS Genetics,* 11**,** e1004999.

FRANCESCONI, S., COPELAND, W., C. & WANG, T., S,F. 1993. *In vivo* species specificity of DNA polymerase α. *Molecular and General Genetics,* 241**,** 457-466.

FREDERICO, L., A., KUNKEL, T., A. & SHAW, B., R. 1990. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry,* 29**,** 2532-2537.

FROMONT-RACINE, M., RAIN, J., C. & LEGRAIN, P. 1997. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genetics,* 16**,** 277-282.

GADALETA, M., C., DAS, M., M., TANIZAWA, H., CHANG, Y., T., NOMA, K., NAKAMURA, T., M. & NOGUCHI, E. 2016. Swi1^{Timeless} prevents repeat instability at fission yeast telomeres. *PLoS ONE,* 12**,** e1005943.

GADALETA, M., C. & NOGUCHI, E. 2017. Regulation of DNA replication through natural impediments in the eukaryotic genome. *Genes,* 8**,** e98.

GAMBUS, A., JONES, R., C., SANCHEZ-DIAZ, A., KANEMALO, M., VAN DUERSEN, F., EDMONDSON, R., D. & LABIB, K. 2006. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nature Cell Biology,* 8**,** 358- 366.

GAMBUS, A., KHOUDOLI, G., A., JONES, R., C. & BLOW, J., J. 2011. MCM2-7 form double hexamers at licensed origins in X*enopus e*gg extract. *The Journal of Biological Chemistry,* 286**,** 11855-11864.

GAMBUS, A., VAN DUERSEN, F., POLYCHRONOPOULOS, D., FOLTMAN, M., JONES, R., C., EDMONDSON, R., D. & LABIB, K. 2009. A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase α within the eukaryotic replisome. *The EMBO Journal,* 28**,** 2992-3004.

GARCES, F., PEARL, L., H. & OLIVER, A., W. 2011. The structural basis for substrate recognition by mammalian polynucleotide kinase 3' phosphatase. *Molecular Cell,* 44**,** 385-396.

GARCÍA-PICHARDO, D., CAÑAS, J., C., GARCÍA-RUBIO, M., L., GÓMEZ-GONZÁLEZ, B., RONDÓN, A., G. & AGUILERA, A. 2017. Histone mutants separate R loop formation from genome instability induction. *Molecular Cell,* 66**,** 597-609.

GASCH, A., P., HUANG, M., METZNER, S., BOTSTEIN, D., ELLEDGE, S., J. & BROWN, P., O. 2001. Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Molecular Biology of the Cell,* 12**,** 2987-3003.

GEORGESCU, R., E., LANGSTON, L., YAO, N., Y., YURIEVA, O., ZHANG, D., FINKELSTEIN, J., AGARWAL, T. & O'DONNELL, M., E. 2014. Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nature Structural and Molecular Biology,* 21**,** 664-670.

GHAZAL, G., GAGNON, J., JACQUES, P., E., LANDRY, J., R., ROBERT, F. & ELELA, S., A. 2009. Yeast RNase III triggers polyadenylation-independent transcription termination. *Molecular Cell,* 36**,** 99-109.

GINNO, P., A., LIM, Y., W., LOTT, P., L., KORF, I. & CHÉDIN, F. 2013. GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Research,* 23**,** 1590-1600.

GINNO, P., A., LOTT, P., L., CHRISTENSEN, H., C., KORF, I. & CHÉDIN, F. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Molecular Cell,* 45**,** 814-825.

GOSNELL, J., A. & CHRISTENSEN, T., W. 2011. Drosophila Ctf4 is essential for efficient DNA replication and normal cell cycle progression. *BMC Molecular Biology,* 12**,** doi: 10.1186/1471-2199-12-13.

GRABCZYK, E., MANCUSO, M. & SAMMARCO, M., C. 2007. A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP *in vitro*. *Nucleic Acids Research,* 35**,** 5351-5359.

GRAF, M., BONETTI, D., LOCKHART, A., SERHAL, K., KELLNER, V., MAICHER, A., JOLIVET, P., TEIXEIRA, M., T. & LUKE, B. 2017. Telomere length determines TERRA and R-loop regulation through the cell cycle. *Cell,* 170**,** 72-85.

GRANDIN, N. & CHARBONNEAU, M. 2007. Mrc1, a non-essential DNA replication protein, is required for telomere end protection following loss of capping by Cdc13, Yku or telomerase. *Molecular Genetics and Genomics,* 277**,** 685–699.

GRANNEMAN, S., KUDLA, G., PETFALSKI, E. & TOLLERVEY, D. 2009. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proceedings for the National Academy of Sciences of the USA,* 106**,** 9613-9618.

GREGER, H. & PROUDFOOT, N., J. 1998. Poly(A) signals control both transcriptional termination and initiation between the tandem *GAL10*

and *GAL7* genes of *Saccharomyces cerevisiae*. *The EMBO Journal,* 17**,** 4771-4779.
GRIFFITHS, A., A., ANDERSEN, P., A. & WAKE, R., G. 1998. Replication terminator protein-based replication fork-arrest systems in various B*acillus s*pecies. *Journal of Bacteriology,* 180**,** 3360-3367.
GROH, M. & GROMAK, N. 2014. Out of balance: R-loops in human disease. *PLoS Genetics,* 10**,** e1004630.
GROH, M., LUFINO, M., M., WADE-MARTINS, R. & GROMAK, N. 2014. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and Fragile X syndrome. *PLoS Genetics,* 10**,** e1004318.
GUAN, C., LI, J., SUN, D., LIU, Y. & LIANG, H. 2017. The structure and polymerase-recognition mechanism of the crucial adaptor protein AND-1 in the human replisome. *The Journal of Biological Chemistry,* 292**,** 9627-9636.
GUENTHER, U., P., HANDOKO, L., LAGGERBAUER, B., JABLONKA, S., CHARI, A., ALZHEIMER, M., OHMER, J., PLÖTTNER, O., GEHRING, N., SICKMANN, A., VON AU, K., SCHUELKE, M. & FISCHER, U. 2009. IGHMBP2 is a ribosome-associated helicase inactive in the neuromuscular disorder distal SMA type 1 (DSMA1). *Human Molecular Genetics,* 18**,** 1288-1300.
GUTZ, H. 1967. "Twin meiosis" and other ambivalences in the life cycle of S*chizosaccharomyces pombe*. *Science,* 158**,** 796-798.
GUTZ, H. & SCHMIDT, H. 1985. Switching genes in S*chizosaccharomyces pombe*. *Current Genetics,* 9**,** 325-331.
HABER, J., E. 2012. Mating-type genes and M*AT s*witching in S*accharomyces cerevisiae*. *Genetics,* 191**,** 33-64.
HABRAKEN, Y., SUNG, P., PRAKASH, L. & PRAKASH, S. 1993. Yeast excision repair gene *RAD2* encodes a single-stranded DNA endonuclease. *Nature,* 366**,** 365-368.
HAMPERL, S., BOCEK, M., J., SALDIVAR, J., C., SWIGUT, T. & CIMPRICH, K., A. 2017. Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell,* 170**,** 774-786.
HANNA, J., S., KROLL, E., S., LUNDBLAD, V. & SPENCER, F., A. 2001. S*accharomyces cerevisiae* C*TF18* and C*TF4* are required for sister chromatid cohesion. *Molecular and Cellular Biology,* 21**,** 3144-3158.
HANSON, S., J. & WOLFE, K., H. 2017. An evolutionary perspective on yeast mating-type switching. *Genetics,* 206**,** 9-32.
HARTWELL, L., H., MORTIMER, R., K., CULOTTI, J. & CULOTTI, M. 1973. Genetic control of the cell division cycle in yeast: V. Genetic analysis of *cdc* mutants. *Genetics,* 74**,** 267–286.
HAYASHI, M., T., TAKAHASHI, T., S., NAKAGAWA, T., NAKAYAM, J. & MASUKATA, H. 2009. The heterochromatin protein Swi6/HP1 activates replication origins at per-centromeric region and silent mating-type locus. *Nature Cell Biology,* 11**,** 357-362.
HAZELBAKER, D., Z., MARQUARDT, S., WLOTZKA, W. & BURATOWSKI, S. 2013. Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. *Molecular Cell,* 49**,** 55-66.

HEICHINGER, C., PENKETT, C., J., BÄHLER, J. & NURSE, P. 2006. Genome-wide characterization of fission yeast DNA replication origins. *The EMBO Journal,* 25**,** 5171–5179.

HELBIG, A., O., ROSATI, S., PIJNAPPEL, P., W., VAN BREUKELEN, B., TIMMERS, M., H., MOHAMMED, S., SLIJPER, M. & HECK, A., J. 2010. Perturbation of the yeast N-acetyltransferase NatB induces elevation of protein phosphorylation levels. *BMC Genomics,* 11**,** 685.

HELLER, R., C., KANG, S., LAM, W., M., CHEN, S., CHAN, C., S. & BELL, S., P. 2011. Eukaryotic origin-dependent DNA replication *in vitro* reveals sequential action of DDK and S-CDK kinases. *Cell,* 146**,** 80-91.

HELMRICH, A., BALLARINO, M., NUDLER, E. & TORA, L. 2013. Transcription-replication encounters, consequences and genomic instability. *Nature Structural and Molecular Biology,* 20**,** 412-418.

HELMRICH, A., BALLARINO, M. & TORA, L. 2011. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Molecular Cell,* 44**,** 966-977.

HEO, D., H., YOO, I., KONG, J., LIDSCHREIBER, M., MAYER, A., CHOI, B., Y., HAHN, Y., CRAMER, P., BURATOWSKI, S. & KIM, M. 2013. The RNA polymerase II C-terminal domain-interacting domain of yeast Nrd1 contributes to the choice of termination pathway and couples to RNA processing by the nuclear exosome. *Journal of Biological Chemistry,* 288**,** 36676-36690.

HERNANDEZ, A., J., LEE, S., J. & RICHARDSON, C., C. 2016. Primer release is the rate-limiting event in lagging strand synthesis mediated by the T7 replisome. *Proceedings for the National Academy of Sciences of the USA,* 113**,** 5916-5921.

HIGGS, D., R., GOODBOURN, S., E., LAMB, J., CLEGG, J., B., WEATHERALL, D., J. & PROUDFOOT, N., J. 1983. Alpha-thalassaemia caused by a polyadenylation signal mutation. *Nature,* 306**,** 398-400.

HODGSON, B., CALZADA, A. & KARIM, L. 2007. Mrc1 and Tof1 regulate DNA replication forks in different ways during normal S phase. *Molecular Biology of the Cell,* 18**,** 3894–3902.

HOGG, M., OSTERMAN, P., BYLUND, G., O., GANAI, R., A., LUNDSTRÖM, E., B., SAUER-ERIKSSON, A., E. & JOHANSSON, E. 2014. Structural basis for processive DNA synthesis by yeast DNA polymerase ε. *Nature Structural and Molecular Biology,* 21**,** 49-55.

HOLMES, A., ROSEAULIN, L., SCHURRA, C., WAXIN, H., LAMBERT, S., ZARATIEGUI, M., MARTIENSSEN, R., A. & ARCANGIOLI, B. 2012. Lsd1 and Lsd2 control programmed replication fork pauses and imprinting in fission yeast. *Cell Reports,* 2**,** 1513-1520.

HSIN, J., P. & MANLEY, J., L. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes and Development,* 26**,** 2119-2137.

HUANG, M., ZHOU, Z. & ELLEDGE, S., J. 1998. The DNA replication and damage checkpoint pathways induce transcription by inhibition of the Crt1 repressor. *Cell,* 94**,** 595-605.

HUERTAS, P. & AGUILERA, A. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular Cell,* 12**,** 711-721.

ILVES, I., PETOJEVIC, T., PESAVENTO, J., J. & BOTCHAN, M., R. 2010. Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Molecular Cell,* 37**,** 247-258.

ITOH, T. & TOMIZAWA, J., I. 1980. Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonucleotide H. *Proceedings for the National Academy of Sciences of the USA,* 77**,** 2450-2454.

IYER, D., R. & RHIND, N. 2017. The intra-S checkpoint responses to DNA damage. *Genes,* 8**,** E74.

JACKSON, B., R., NOERENBERG, M. & WHITEHOUSE, A. 2014. A novel mechanism inducing genome instability in Kaposi's sarcoma-associated herpesvirus infected cells. *PLoS Pathology,* 10**,** e1004098.

JAKOČIŪNAS, T., HOLM, L., R., VERHEIN-HANSEN, J., TRUSINA, A. & THON, G. 2013. Two portable recombination enhancers direct donor choice in fission yeast heterochromatin. *PLoS Genetics,* 9**,** e1003762.

JAMONNAK, N., CREAMER, T., J., DARBY, M., M., SCHAUGHENCY, P., WHEELAN, S., J. & CORDEN, J., L. 2011. Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. *RNA,* 17**,** 2011-2025.

JANKE, C., MAGIERA, M., M., RATHFELDER, N., TAXIS, C., REBER, S., MAEKAWA, H., MORENO-BORCHART, A., DOENGES, G., SCHWOB, E., SCHIEBEL, E. & KNOP, M. 2004. A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast,* 21**,** 947-962.

JANKOWSKY, E. 2011. RNA helicases at working: binding and rearranging. *Trends in Biochemical Sciences,* 36**,** 19-29.

JANSEN, L., E,T., DEN BULK, H., BROUNS, R., M., DE RUIJTER, M., BRANDSMA, J., A. & BROUWER, J. 2000. Spt4 modulates Rad26 requirement in transcription-coupled nucleotide excision repair. *The EMBO Journal,* 19**,** 6498–6507.

JOHANSSON, E. & DIXON , N. 2013. Replicative DNA polymerases. *Cold Spring Harbor Perspectives in Biology,* 5**,** a012799.

JOHNSON, R., E., KLASSEN, R., PRAKASH, L. & PRAKASH, S. 2015. A major role of DNA Polymerase δ in replication of both the leading and lagging DNA strands. *Molecular Cell,* 59**,** 163-175.

JOYCE, C., M. 1997. Choosing the right sugar: How polymerases select a nucleotide substrate. *Proceedings for the National Academy of Sciences of the USA,* 94**,** 1619–1622.

KAGUNI, L., S. & CLAYTON, D., A. 1982. Template-directed pausing in *in vitro* DNA synthesis by DNA polymerase α from D*rosophila melanogaster* embryos. *Proceedings for the National Academy of Sciences of the USA,* 79 983-987.

KAMADA, K., HORIUCHI, T., OHSUMI, K., SHIMAMOTO, N. & MORIKAWA, K. 1996. Structure of a replication-terminator protein complexed with DNA. *Nature,* 383**,** 598-603.

KAMIMURA, Y., TAK, Y., S., SUGINO, A. & ARAKI, H. 2001. Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in S*accharomyces cerevisiae*. *The EMBO Journal,* 20**,** 2097-2107.

KANEMAKI, M., T. & LABIB, K. 2006. Distinct roles for Sld3 and GINS during establishment and progression of eukaryotic DNA replication forks. *The EMBO Journal,* 25**,** 1753–1763.

KANKE, M., KODAMA, Y., TAKAHASHI, T., S., NAKAGAWA, T. & MASUKATA, H. 2012. Mcm10 plays an essential role in origin DNA unwinding after loading of the CMG components. *The EMBO Journal,* 31**,** 2182-2194.

KANKE, M., NISHIMURA, K., KANEMAKI, M., T., KAKIMOTO, T., TAKAHASHI, T., NAKAGAWA, T. & MASUKATA, H. 2011. Auxin-inducible protein depletion system in fission yeast. *BMC Cell Biology,* 12**,** 8.

KAWAUCHI, J., MISCHO, H., BRAGLIA, P., RONDON, A. & PROUDFOOT, N., J. 2008. Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes and Development,* 22**,** 1082-1092.

KAYKOV, A. & ARCANGIOLI, B. 2004. A programmed strand-specific and modified nick in S*. pombe* constitutes a novel type of chromosomal imprint. *Current Biology,* 14**,** 1924-1928.

KELLEY, L., A., MEZULIS, S., YATES, C., M., WASS, M., N. & STERNBERG, M., J. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols,* 10**,** 845-858

KESKIN, H., SHEN, Y., HUANG, F., PATEL, M., YANG, T., ASHLEY, K., MAZIN, A., V. & STORICI, F. 2014. Transcript-RNA-templated DNA recombination and repair. *Nature,* 515**,** 436-439.

KESZTHELYI, A., DAIGAKU, Y., PTASIŃSKA, K., MIYABE, I. & CARR, A., M. 2015. Mapping ribonucleotides in genomic DNA and exploring replication dynamics by polymerase usage sequencing (Pu-seq). *Nature Protocols,* 10**,** 1786-1801.

KILCHERT, C., WITTMANN, S. & VASILJEVA, L. 2016. The regulation and functions of the nuclear RNA exosome complex. *Nature Reviews in Molecular and Cellular Biology,* 17**,** 227-239.

KILKENNY, M., L., DE PICCOLI, G., PERERA, R., L., LABIB, K. & PELLEGRINI, L. 2012. A conserved motif in the C-terminal tail of DNA polymerase α tethers primase to the eukaryotic replisome. *The Journal of Biological Chemistry,* 287**,** 23740-23747.

KIM, H., D., CHOE, J. & SEO, Y., S. 1999. The s*en1*$^{+}$gene of S*chizosaccharomyces pombe,* a homologue of budding yeast S*EN1,* encodes an RNA and DNA helicase. *Biochemistry,* 38**,** 14697-14710.

KIM, K., HEO, D., H., KIM, I., SUH, J., Y. & KIM, M. 2016. Exosome cofactors connect transcription termination to RNA processing by guiding terminated transcripts to the appropriate exonuclease within the nuclear exosome. *Journal of Biological Chemistry,* 291**,** 13229-13242.

KIM, M., KROGAN, N., J., VASILJEVA, L., RANDO, O., J., NEDEA, E., GREENBLATT, J., F. & BURATOWSKI, S. 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature,* 432**,** 517-522.

KIM, M., VASILJEVA, L., RANDO, O., J., ZHELKOVSKY, A., MOORE, C. & BURATOWSKI, S. 2006. Distinct pathways for snoRNA and mRNA termination. *Molecular Cell,* 24**,** 723-734.

KIM, N., ABDULOVIC, A., L., GEALY, R., LIPPERT, M., J. & JINKS-ROBERTSON, S. 2007. Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of DNA replication. *DNA Repair,* 6**,** 1285-1296.

KIM, S., M., DUBEY, D., D. & HUBERMAN, J., A. 2003. Early-replicating heterochromatin. *Genes and Development,* 17**,** 330-335.

KIPREOS, E., T. 2005. C*. elegans* cell cycles: invariance and stem cell divisions. *Nature Reviews in Molecular and Cellular Biology,* 6**,** 766-776.

KLAR, A., J,S. & MIGLIO, L., M. 1986. Initiation of meiotic recombination by double-strand DNA breaks in S*. pombe*. *Cell,* 46**,** 725-731.

KOBAYASHI, T. 2003. The replication fork barrier site forms a unique structure with Fob1p and inhibits the replication fork. *Molecular and Cellular Biology,* 23**,** 9178–9188.

KOBAYASHI, T. & HORIUCHI, T. 1996. A yeast gene product, Fob1 protein, required for both replication fork blocking and recombinational hotspot activities. *Genes to Cells,* 1**,** 465-474.

KOGOMA, T. 1997. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiology and Molecular Biology Reviews,* 61**,** 212-238.

KOGOMA, T. & VON MEYENBURG, K. 1983. The origin of replication, o*riC,* and the d*naA* protein are dispensable in stable DNA replication (s*drA)* mutants of E*scherichia coli* K-12. *The EMBO Journal,* 2**,** 463-468.

KOULINTCHENKO, M., VENGROVA, S., EYDMANN, T., ARUMUGAM, P. & DALGAARD, J., Z. 2012. DNA Polymerase (s*wi7)* and the flap endonuclease Fen1 (r*ad2)* act together in the S-phase alkylation damage response in S*. pombe*. *PLoS ONE,* 7**,** e47091.

KOUPRINA, N., KROLL, E., BANNIKOV, V., BLISKOVSKY, V., GIZATULLIN, R., KIRILLOV, A., SHESTOPALOV, B., ZAKHARYEV, V., HIETER, P., SPENCER, F. & LARIONOV, V. 1992. *CTF4* (*CHL15*) mutants exhibit defective DNA metabolism in the yeast *Saccharomyces cerevisiae*. *Molecular and Cellular Biology,* 12**,** 5736-5747.

KRASTANOVA, I., SANNINO, V., AMENITSCH, H., GILEADI, O., PISANI, F. M. & ONESTI, S. 2012. Structural and functional insights into the DNA replication factor Cdc45 reveal an evolutionary relationship to the DHH family of phosphoesterases. *The Journal of Biological Chemistry,* 287**,** 4121–4128.

KRINGS, G. & BASTIA, D. 2004. s*wi1-* and s*wi3*-dependent and independent replication fork arrest at the ribosomal DNA of S*chizosaccharomyces pombe*. *Proceedings for the National Academy of Sciences of the USA,* 101**,** 14085-14090.

KRINGS, G. & BASTIA, D. 2005. Sap1p binds to Ter1 at the ribosomal DNA of S*chizosaccharomyces pombe* and causes polar replication fork arrest. *The Journal of Biological Chemistry,* 280**,** 39135-39142.

KRISHNA, T., S., KONG, X., P., GARY, S., BURGERS, P., M. & KURIYAN, J. 1994. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell,* 79**,** 1233-1243.

KUNKEL, T., A. & BURGERS, P., M. 2008. Dividing the workload at a eukaryotic replication fork. *Trends in Cell Biology,* 18**,** 521-527.

KUNST, F., OGASAWARA, N., MOSZER, I., ALBERTINI, A., M., ALLONI, G., AZEVEDO, V., BERTERO, M., G., BESSIÈRES, P., BOLOTIN, A., BORCHERT, S., BORRISS, R., BOURSIER, L., BRANS, A., BRAUN, M., BRIGNELL, S., C., BRON, S., BROUILLET, S., BRUSCHI, C., V., CALDWELL, B., CAPUANO, V., CARTER, N., M., CHOI, S., K.,

CORDANI, J., J., CONNERTON, I., F., CUMMINGS, N., J., DANIEL, R., A., DENZIOT, F., DEVINE, K., M., DÜSTERHÖFT, A., EHRLICH, S., D., EMMERSON, P., T., ENTIAN, K., D., ERRINGTON, J., FABRET, C., FERRARI, E., FOULGER, D., FRITZ, C., FUJITA, M., FUJITA, Y., FUMA, S., GALIZZI, A., GALLERON, N., GHIM, S., Y., GLASER, P., GOFFEAU, A., GOLIGHTLY, E., J., GRANDI, G., GUISEPPI, G., GUY, B., J., HAGA, K., HAIECH, J., HARWOOD, C., R., HÈNAUT, A., HILBERT, H., HOLSAPPEL, S., HOSONO, S., HULLO, M., F., ITAYA, M., JONES, L., JORIS, B., KARAMATA, D., KASAHARA, Y., KLAERR-BLANCHARD, M., KLEIN, C., KOBAYASHI, Y., KOETTER, P., KONINGSTEIN, G., KROGH, S., KUMANO, M., KURITA, K., LAPIDUS, A., LARDINOIS, S., LAUBER, J., LAZAREVIC, V., LEE, S., M., LEVINE, A., LIU, H., MASUDA, S., MAUËL, C., MÉDIGUE, C., MEDINA, N., MELLADO, R., P., MIZUNO, M., MOESTL, D., NAKAI, S., NOBACK, M., NOONE, D., O'REILLY, M., OGAWA, K., OGIWARA, A., OUDEGA, B., PARK, S., H., PARRO, V., POHL, T., M., PORTELLE, D., PORWOLLIK, S., PRESCOTT, A., M., PRESECAN, E., PUJIC, P., PURNELLE, B., et al. 1997. The complete genome sequence of the gram-positive bacterium B*acillus subtilis*. *Nature,* 390**,** 249-256.

LABIB, K. 2010. How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? *Genes and Development,* 24**,** 1208-1219.

LABIB, K. & HODGSON, B. 2007. Replication fork barriers: pausing for a break or stalling for time? *The EMBO Reports,* 8**,** 346-353.

LAMBERT, S., WATSON, A., SHEEDY, D., M., MARTIN, B. & CARR, A., M. 2005. Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier. *Cell,* 121**,** 689-702.

LANGSTON, L., ZHANG, D., YURIEVA, O., GEORGESCU, R., E., FINKELSTEIN, J., YAO, N., INDIANI, C. & O'DONNELL, M., E. 2014. CMG helicase and DNA polymerase ε form a functional 15-subunit holoenzyme for eukaryotic leading-strand replication. *Proceedings for the National Academy of Sciences of the USA,* 111**,** 15390-15395.

LAPIDOT , A., BARAN , N. & MANOR, H. 1989. (dT-dC)n and (dG-dA)n tracts arrest single stranded DNA replication i*n vitro*. *Nucleic Acids Research,* 17**,** 883–900.

LARSEN, N., B., SASS, E., SUSKI, C., MANKOURI, H., W. & HICKSON, I., D. 2014. The E*scherichia coli T*us-Ter replication fork barrier causes site-specific DNA replication perturbation in yeast. *Nature Communications,* 5.

LAZZARO, F., NOVARINA, D., AMARA, F., WATT, D., L., STONE, J., E., COSTANZO, V., BURGERS, P., M., KUNKEL, T., A., PLEVANI, P. & MUZI-FALCONI, M. 2012. RNase H and postreplication repair protect cells from ribonucleotides incorporated in DNA. *Molecular Cell,* 45**,** 99-110.

LE BER, I., BOUSLAM, N., RIVAUD-PÉCHOUX, S., GUIMARÃES, J., BENOMAR, A., CHAMAYOU, C., GOIZET, C., MOREIRA, M., C., KLUR, S., YAHYAOUI, M., AGID, Y., KOENIG, M., STEVANIN, G., BRICE, A. & DÜRR, A. 2004. Frequency and phenotypic spectrum of

ataxia with oculomotor apraxia 2: a clinical and genetic study in 18 patients. *Brain,* 127**,** 759-767.
LEE, J., B., HITE, R., K., HAMDAN, S., M., XIE, X., S., RICHARDSON, C., C. & VAN OIJEN, A., M. 2006. DNA primase acts as a molecular brake in DNA replication. *Nature,* 439**,** 621-624.
LEGROS, P., MALAPERT, A., NIINUMA, S., BERNARD, P. & VANOOSTHUYSE, V. 2014. RNA processing factors Swd2.2 and Sen1 antagonize RNA Pol III-dependent transcription and the localization of condensin at Pol III genes. *PLoS Genetics,* 10**,** e1004794.
LEONAITĖ, B., HAN, Z., BASQUIN, J., BONNEAU, F., LIBRI, D., PORRUA, O. & CONTI, E. 2017. Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family. *The EMBO Journal,* 36**,** 1590-1604.
LESNIK, E., A. & FREIER, S., M. 1995. Relative thermodynamic stability of DNA, RNA, and DNA: RNA hybrid duplexes: Relationship with base composition and structure. *Biochemistry,* 34**,** 10807-10815.
LI, L., B., YU, Z., TENG, X. & BONINI, N., M. 2008. RNA toxicity is a component of ataxin-3 degeneration in D*rosophila*. *Nature,* 453**,** 1107-1111.
LI, P., C., CHRETIEN, L., CÔTÉ, J., KELLY, T., J. & FORSBURG, S., L. 2011. *S. pombe* replication protein Cdc18 (Cdc6) interacts with Swi6 (HP1) heterochromatin protein: region specific effects and replication timing in the centromere. *Cell Cycle,* 10**,** 323-336.
LI, S. & SMERDON, M., J. 2002. Rpb4 and Rpb9 mediate subpathways of transcription-coupled DNA repair in S*accharomyces cerevisiae*. *The EMBO Journal,* 21**,** 5921–5929.
LI, W., COWLEY, A., ULUDAG, M., GUR, T., MCWILLIAM, H., SQUIZZATO, S., PARK, Y., M., BUSO, N. & LOPEZ, R. 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research,* 43**,** W580-W584.
LI, W., SELVAM, K., RAHMAN, S., A. & LI, S. 2016. Sen1, the yeast homolog of human senataxin, plays a more direct role than Rad26 in transcription coupled DNA repair. *Nucleic Acids Research,* 44**,** 6794–6802.
LI, X. & MANLEY, J., L. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell,* 122**,** 365-378.
LIANG, C. & STILLMAN, B. 1997. Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in c*dc6* mutants. *Genes and Development,* 11**,** 3375-3386.
LISBY, M., ROTHSTEIN, R. & MORTENSEN, U., H. 2001. Rad52 forms DNA repair and recombination centers during S phase. *Proceedings for the National Academy of Sciences of the USA,* 98**,** 8276-8282.
LITTLE, R., D., PLATT, T., H. & SCHILDKRAUT, C., L. 1993. Initiation and termination of DNA replication in human rRNA genes. *Molecular and Cellular Biology,* 13**,** 6600-6613.
LIU, B. & ALBERTS, B. 1995. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science,* 267**,** 1131-1137.
LOPES, M., COTTA-RAMUSINO, C., PELLICIOLI, A., LIBERI, G., PLEVANI, P., MUZI-FALCONI, M., NEWLON, C., S. & FOIANI, M. 2001. The DNA

replication checkpoint response stabilizes stalled replication forks. *Nature,* 412**,** 557-561.

LOPES, M., FOIANI, M. & SOGO, J., M. 2006. Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. *Molecular Cell,* 21**,** 15-27.

LOPEZ-MOSQUEDA, J., MAAS, N., L., JONSSON, Z., O., DEFAZIO-ELI, L., G., WOHLSCHLEGEL, J. & TOCZYSKI, D., P. 2010. Damage-induced phosphorylation of Sld3 is important to block late origin firing. *Nature,* 467**,** 479-483.

LUKE, B., PANZA, A., REDON, S., IGLESIAS, N., LI, Z. & LINGNER, J. 2008. The Rat1p 5' to 3' exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in S*accharomyces cerevisiae*. *Molecular Cell,* 32**,** 465-477.

LUNDBLAD, V. & BLACKBURN, E., H. 1993. An alternative pathway for yeast telomere maintenance rescues e*st-* senescence. *Cell,* 73**,** 347-360.

LÓPEZ-ESTRAÑO, C., SCHVARTZMAN, J., B., KRIMER, D., B. & HERNÁNDEZ, P. 1999. Characterization of the pea rDNA replication fork barrier: putative c*is*-acting and t*rans*-acting factors. *Plant Molecular Biology,* 40**,** 99-110.

LÕOKE, M., MALONEY, M., F. & BELL, S., P. 2017. Mcm10 regulates DNA replication elongation by stimulating the CMG replicative helicase. *Genes and Development,* 31**,** 291-305.

MACALPINE, D., M., ZHANG, Z. & KAPLER, G., M. 1997. Type I elements mediate replication fork pausing at conserved upstream sites in the *Tetrahymena thermophila* ribosomal DNA minichromosome. *Molecular and Cellular Biology,* 17**,** 4517-4525.

MAJKA, J. & BURGERS, P., M. 2004. The PCNA-RFC families of DNA clamps and clamp loaders. *Progress in Nucleic Acid Research and Molecular Biology,* 78**,** 227-260.

MAMNUN, Y., A., TAKAYAMA, S. & TODA, T. 2006. Fission yeast Mcl1 interacts with SCFP^{of3} and is required for centromere formation. *Biochemical and Biophysical Research Communications,* 350**,** 125-130.

MANNERS, D., J. & MEYER, M., T. 1977. The molecular structures of some glucans from the cell walls of S*chizosaccharomyces pombe*. *Carbohydrate Research,* 57**,** 189-203.

MAO, P., SMERDON, M., J., ROBERTS, S., A. & WYRICK, J., J. 2016. Chromosomal landscape of UV damage formation and repair at single-nucleotide resolution. *Proceedings for the National Academy of Sciences of the USA,* 113**,** 9057–9062.

MARIC, C., LEVACHER, B. & HYRIEN, O. 1999. Developmental regulation of replication fork pausing in *Xenopus laevis* ribosomal RNA genes. *Journal of Molecular Biology,* 291**,** 775-788.

MARIC, M., MACULINS, T., DE PICCOLI, G. & LABIB, K. 2014. Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. *Science,* 346**,** 1253596.

MARTEIJN, J., A., LANS, H., VERMEULEN, W. & HOEIJMAKERS, J., H,J. 2014. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature Reviews Molecular Cell Biology,* 15**,** 465–481.

MARTIN-TUMASZ, S. & BROW, D., A. 2015. *Saccharomyces cerevisiae* Sen1 helicase domain exhibits 5'- to 3'- helicase activity with a preference for translocation on DNA rather than RNA. *Journal of Biological Chemistry,* 290**,** 22880-22889.

MCCRACKEN, S., FONG, N., YANKULOV, K., BALLANTYNE, S., PAN, G., GREENBLATT, J., PATTERSON, S., D., WICKENS, M. & BENTLEY, D., L. 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature,* 385**,** 357-361.

MENDENHALL , M., D. & HODGE , A., E. 1998. Regulation of Cdc28 cyclin-dependent protein kinase activity during the cell cycle of the yeast *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews,* 62**,** 1191–1243.

MILES, J. & FORMOSA, T. 1992a. Evidence that P*OB1,* a S*accharomyces cerevisiae* protein that binds to DNA polymerase alpha, acts in DNA metabolism *in vivo*. *Molecular and Cellular Biology,* 12**,** 5724-5735.

MILES, J. & FORMOSA, T. 1992b. Protein affinity chromatography with purified yeast DNA polymerase alpha detects proteins that bind to DNA polymerase. *Proceedings for the National Academy of Sciences of the USA,* 89**,** 1276-1280.

MILLER, M., S., RIALDI, A., HO, J., S., TILOVE, M., MARTINEZ-GIL, L., MOSHKINA, N., P., PERALTA, Z., NOEL, J., MELEGARI, C., MAESTRE, A., M., MITSOPOULOS, P., MADRENAS, J., HEINZ, S., BENNER, C., YOUNG, J., A., FEAGINS, A., R., BASLER, C., F., FERNANDEZ-SESMA, A., BECHEREL, O., J., LAVIN, M., F., VAN BAKEL, H. & MARAZZI, I. 2015. The helicase senataxin suppresses the antiviral transcriptional response and controls viral biogenesis. *Nature Immunology,* 16**,** 485–494.

MINNICK, D., T., ASTATKE, M., JOYCE, C., M. & KUNKEL, T., A. 1996. A thumb subdomain mutant of the large fragment of *Escherichia coli* DNA Polymerase I with reduced DNA binding affinity, processivity, and frameshift fidelity. *The Journal of Biological Chemistry,* 271**,** 24954-24961.

MIRKIN, E., V. & MIRKIN, S., M. 2007. Replication fork stalling at natural impediments. *Microbiology and Molecular Biology Reviews,* 71**,** 13-35.

MISCHO, H., E., GÓMEZ-GONZÁLEZ, B., GRZECHNIK, P., RONDÓN, A., G., WEI, W., STEINMETZ, L., AGUILERA, A. & PROUDFOOT, N., J. 2011. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Molecular Cell,* 41**,** 21-32.

MITSUI, J., TAKAHASHI, Y., GOTO, J., TOMIYAMA, H., ISHIKAWA, S., YOSHINO, H., MINAMI, N., SMITH, D., I., LESAGE, S., ABURATANI, H., NISHINO, I., BRICE, A., HATTORI, N. & TSUJI, S. 2010. Mechanisms of genomic instabilities underlying two common fragile-site-associated loci, PARK2 and DMD, in germ cell and cancer cell lines. *American Journal of Human Genetics,* 87**,** 75-89.

MIYABE, I., KUNKEL, T., A. & CARR, A., M. 2011. The major roles of DNA polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved. *PLoS Genetics,* 7**,** e1002407.

MIYATA, H. & MIYATA, M., J. 1981. Mode of conjugation in homothallic cells of S*chizosaccharomyces pombe*. *The Journal of General and Applied Microbiology,* 27**,** 365-371.

MOHANTY, B., K. & BASTIA, D. 2004. Binding of the replication terminator protein Fob1p to the Ter sites of yeast causes polar fork arrest. *The Journal of Biological Chemistry,* 279**,** 1932-1941.

MOREIRA, M., C., KLUR, S., WATANABE, M., NÉMETH, A., H., LE BER, I., MONIZ, J., C., TRANCHANT, C., AUBOURG, P., TAZIR, M., SCHÖLS, L., PANDOLFO, M., SCHULZ, J., B., POUGET, J., CALVAS, P., SHIZUKA-IKEDA, M., SHOJI, M., TANAKA, M., IZATT, L., SHAW, C., E., M'ZAHEM, A., DUNNE, E., BOMONT, P., BENHASSINE, T., BOUSLAM, N., STEVANIN, G., BRICE, A., GUIMARÃES, J., MENDONÇA, P., BARBOT, C., COUTINHO, P., SEQUEIROS, J., DÜRR, A., WARTER, J., M. & KOENIG, M. 2004. Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nature Genetics,* 36**,** 225-227.

MULCAIR, M., D., SCHAEFFER, P., M., OAKLEY, A., J., CROSS, H., F., NEYLON, C., HILL, T., M. & DIXON, N., E. 2006. A molecular mousetrap determines polarity of termination of DNA replication in *E. coli*. *Cell,* 125**,** 1309-1319.

MURAKAMI, H. & OKAYAMA, H. 1995. A kinase from fission yeast responsible for blocking mitosis in S phase. *Nature,* 374**,** 817-819.

NAKAYAMA, J., ALLSHIRE, R., C., KLAR, A., J,S. & GREWAL, S., I,S. 2001a. A role for DNA polymerase α in epigenetic control of transcriptional silencing in fission yeast. *The EMBO Journal,* 20**,** 2857-2866.

NAKAYAMA, J., KLAR, A., J,S. & GREWAL, S., I,S. 2000. A chromodomain protein, Swi6, perform imprinting functions in fission yeast during mitosis and meiosis. *Cell,* 101**,** 307-317.

NAKAYAMA, J., RICE, J., C., STRAHL, B., D., ALLIS, C., D. & GREWAL, S., I,S. 2001b. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science,* 292**,** 110-113.

NEDEA, E., HE, X., KIM, M., POOTOOLAL, J., ZHONG, G., CANADIEN, V., HUGHES, T., BURATOWSKI, S., MOORE, C., L. & GREENBLATT, J. 2003. Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. *Journal of Biological Chemistry,* 278**,** 33000-33010.

NEDEA, E., NALBANT, D., XIA, D., THEOHARIS, N., T., SUTER, B., RICHARDSON, C., J., TATCHELL, K., KISLINGER, T., GREENBLATT, J., F. & NAGY, P., L. 2008. The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Molecular Cell,* 29**,** 577-587.

NETZ, D., J., STITH, C., M., STÜMPFIG, M., KÖPF, G., VOGEL, D., GENAU, H., M., STODOLA, J., L., LILL, R., BURGERS, P., M. & PIERIK, A., J. 2011. Eukaryotic DNA polymerases require an iron-sulfur cluster for the formation of active complexes *Nature Chemical Biology,* 8**,** 125-132.

NEYLON, C., BROWN, S., E., KRALICEK, A., V., MILES, C., S., LOVE, C., A. & DIXON, N., E. 2000. Interaction of the E*scherichia coli* replication terminator protein (Tus) with DNA: A model derived from DNA-binding studies of mutant proteins by surface plasmon resonance. *Biochemistry,* 39**,** 11989-11999.

NICK MCELHINNY, S., A., KUMAR, D., CLARK, A., B., WATT, D., L., WATTS, B., E., LUNDSTRÖM, E., B., JOHANSSON, E., CHABES, A. & KUNKEL, T., A. 2010a. Genome instability due to ribonucleotide incorporation into DNA. *Nature Chemical Biology,* 6**,** 774-781.

NICK MCELHINNY, S., A., WATTS, B., E., KUMAR, D., WATT, D., L., LUNDSTRÖM, E., B., BURGERS, P., M., JOHANSSON, E., CHABES, A. & KUNKEL, T., A. 2010b. Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proceedings for the National Academy of Sciences of the USA,* 107**,** 4949-4954.

NIELSEN, S., YUZENKOVA, Y. & ZENKIN, N. 2013. Mechanism of eukaryotic RNA polymerase III transcription termination. *Science,* 340**,** 1577-1580.

NOGUCHI, E., NOGUCHI, C., DU, L., L. & RUSSELL, P. 2003. Swi1 prevents replication fork collapse and controls checkpoint kinase Cds1. *Molecular and Cellular Biology,* 23**,** 7861-7874.

NOGUCHI, E., NOGUCHI, C., MCDONALD, W., H., YATES, J., R,3RD. & RUSSELL, P. 2004. Swi1 and Swi3 are components of a replication fork protection complex in fission yeast. *Molecular and Cellular Biology,* 24**,** 8342-8355.

NOJIMA, T., GOMES, T., GROSSO, A., R., KIMURA, H., DYE, M., J., DHIR, S., CARMO-FONSECA, M. & PROUDFOOT, N., J. 2015. Mammalian NET-Seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell,* 161**,** 526-540.

NURSE, P. & BISSETT, Y. 1981. Gene required in G_1 for commitment to cell cycle and in G_2 for control of mitosis in fission yeast. *Nature,* 292**,** 558-5560.

O'CONNELL, K., JINKS-ROBERTSON, S. & PETES, T., D. 2015. Elevated genome-wide instability in yeast mutants lacking RNase H activity. *Genetics,* 201**,** 963-975.

OBARA-ISHIHARA, T. & OKAYAMA, H. 1994. A B-type cyclin negatively regulates conjugation via interacting with cell cycle 'start' genes in fission yeast. *The EMBO Journal,* 13**,** 1863-1872.

OLOVNIKOV, A., M. 1996. Telomeres, telomerase, and aging: Origin of the theory. *Experimental Gerontology,* 31**,** 443-448.

ORKIN, S., H., CHENG, T., C., ANTONARAKIS, S., E. & KAZAZIAN, H., H,JR. 1985. Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. *The EMBO Journal,* 4**,** 453-456.

OROZCO, I., J., KIM, S., J. & MARTINSON, H., G. 2002. The poly(A) Signal, without the assistance of any downstream element, directs RNA Polymerase II to pause i*n vivo* and then to release stochastically from the template. *Journal of Biological Chemistry,* 277**,** 42899-42911.

ORTIZ-BAZÁN, M., Á., GALLO-FERNÁNDEZ, M., SAUGAR, I., JIMÉNEZ-MARTÍN, A., VÁZQUEZ, M., V. & TERCERO, J., A. 2014. Rad5 plays a major role in the cellular response to DNA damage during chromosome replication. *Cell Reports,* 9**,** 460-468.

OSBORN, A., J. & ELLEDGE, S., J. 2003. Mrc1 is a replication fork component whose phosphorylation in response to DNA replication stress activates Rad53. *Genes and Development,* 17**,** 1755-1767.

PARK, H. & STERNGLANZ, R. 1999. Identification and characterization of the genes for two topoisomerase I-interacting proteins from S*accharomyces cerevisiae*. *Yeast,* 15**,** 35-41.

PARK, J., KANG, M. & KIM, M. 2015. Unraveling the mechanistic features of RNA polymerase II termination by the 5'-3' exoribonuclease Rat1. *Nucleic Acids Research,* 43**,** 2625-2637.

PATEL, P., K., ARCANGIOLI, B., BAKER, S., P., BENSIMON, A. & RHIND, N. 2006. DNA replication origins fire stochastically in fission yeast. *Molecular Biology of the Cell,* 17**,** 308-316.

PATO, M., L. 1975. Alterations of the rate of movement of deoxyribonucleic acid replication forks. *Journal of Bacteriology,* 123**,** 272-277.

PATTURAJAN, M., WEI, X., BEREZNEY, R. & CORDEN, J., L. 1998. A nuclear matrix protein interacts with the phosphorylated C-terminal domain of RNA polymerase II. *Molecular and Cellular Biology,* 18**,** 2406-2415.

PAULSEN, R., D., SONI, D., V., WOLLMAN, R., HAHN, A., T., YEE, M., C., GUAN, A., HESLEY, J., A., MILLER, S., C., CROMWELL, E., F., SOLOW-CORDERO, D., E., MEYER, T. & CIMPRICH, K., A. 2009. A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. *Molecular Cell,* 35**,** 228-239.

PERERA, R., L., TORELLA, R., KLINGE, S., KILKENNY, M., L., MAMAN, J., D. & PELLEGRINI, L. 2013. Mechanism for priming DNA synthesis by yeast DNA Polymerase a. *eLife,* 2**,** e00482.

PETERS, J., M. 2006. The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nature Reviews Molecular Cell Biology,* 7**,** 644-656.

PETOJEVIC, T., PESAVENTO, J., J., COSTA, A., LIANG, J., WANG, Z., BERGER, J., M. & BOTCHAN, M., R. 2015. Cdc45 (cell division cycle protein 45) guards the gate of the eukaryote replisome helicase stabilizing leading strand engagement. *Proceedings for the National Academy of Sciences of the USA,* 112**,** E249-E258.

PFEIFFER, V., CRITTIN, J., GROLIMUND, L. & LINGNER, J. 2013. The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. *The EMBO Journal,* 32**,** 2861-2871.

PFLEIDERER, C., SMID, A., BARTSCH, I. & GRUMMT, I. 1990. An undecamer DNA sequence directs termination of human ribosomal gene transcription. *Nucleic Acids Research,* 18**,** 4727-4736.

POHJOISMÄKI, J., L,O., HOLMES, J., B., WOOD, S., R., YANG, M., Y., YASUKAWA, T., REYES, A., LAURA, J., B., CLUETT, T., J., GOFFART, S., WILLCOX, S., RIGBY, R., E., JACKSON, A., P., SPELBRINK, J., N., GRIFFITH, J., D., CROUCH, R., J., JACOBS, H., T. & HOLT, I., J. 2010. Mammalian mitochondrial DNA replication intermediates are essentially duplex, but contain extensive tracts of RNA/DNA hybrid. *Journal of Molecular Biology,* 397**,** 1144-1155.

PORRUA, O. & LIBRI, D. 2013. A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nature Structural and Molecular Biology,* 20**,** 884-891.

POWELL, W., T., COULSON, R., L., GONZALES, M., L., CRARY, F., K., WONG, S., S., ADAMS, S., ACH, R., A., TSANG, P., YAMADA, N., A., YASUI, D., H., CHÉDIN, F. & LASALLE, J., M. 2013. R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proceedings for the National Academy of Sciences of the USA,* 110**,** 13938-13943.

PRADO, F. & AGUILERA, A. 2005. Impairment of replication fork progression mediates RNA Pol II transcription-associated recombination. *The EMBO Journal,* 24**,** 1267-1276.
PROUDFOOT, N., J. 2011. Ending the message: poly(A) signals then and now. *Genes and Development,* 25**,** 1770-1782.
PROUDFOOT, N., J. 2016. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science,* 352**,** aad9926.
RAO, P., N. & JOHNSON, R., T. 1970. Mammalian cell fusion: studies on the regulation of DNA synthesis and mitosis. *Nature,* 225**,** 159-164.
RARAN-KURUSSI, S. & WAUGH, D., S. 2012. The ability to enhance the solubility of its fusion partners is an intrinsic property of Maltose-Binding Protein but their folding is either spontaneous or chaperone-mediated. *PLoS ONE,* 7**,** e49589.
REDON, C., PILCH, D., R. & BONNER, W., M. 2006. Genetic analysis of S*accharomyces cerevisiae* H2A serine 129 mutant suggests a functional relationship between H2A and the sister-chromatid cohesion partners Csm3-Tof1 for the repair of topoisomerase I-induced DNA damage. *Genetics,* 172**,** 67-76.
REMUS, D., BEURON, F., TOLUN, G., GRIFFITH, J., D., MORRIS, E., P. & DIFFLEY, J., F. 2009. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell,* 139**,** 719-730.
RICHARD, P., FENG, S. & MANLEY, J., L. 2013. A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage. *Genes and Development,* 27**,** 2227-2232.
RICHARD, P. & MANLEY, J., L. 2009. Transcription termination by nuclear RNA. *Genes and Development,* 23**,** 1247-1269.
RIDDLE, M., R., SPICKARD, E., A., JEVINCE, A., NGUYEN, K., C,Q., HALL, D., H., JOSHI, P., M. & ROTHMAN, J., H. 2017. Transorganogenesis and transdifferentiation in C*. elegans* are dependent on differentiated cell identity. *Developmental Biology,* 420**,** 136-147.
ROBERTS, R., W. & CROTHERS, D., M. 1992. Stability and properties of double and triple helices: Dramatic effects of RNA or DNA backbone composition. *Science,* 258**,** 1463-1466.
ROSSI, M., L. & BAMBARA, R., A. 2006. Reconstituted Okazaki fragment processing indicates two pathways of primer removal. *The Journal of Biological Chemistry,* 281**,** 26051-26061.
ROY, D. & LIEBER, M., R. 2009. G clustering is important for the initiation of transcription-induced R-loops i*n vitro,* whereas high G density without clustering is sufficient thereafter. *Molecular and Cellular Biology,* 29**,** 3124-3133.
ROY, D., YU, K. & LIEBER, M., R. 2008. Mechanism of R-loop formation at immunoglobulin class switch sequences. *Molecular and Cellular Biology,* 28**,** 50-60.
ROY, D., ZHANG, Z., LU, Z., HSIEH, C., L. & LIEBER, M., R. 2010. Competition between the RNA transcript and the nontemplate DNA Strand during R-Loop formation i*n vitro:* a nick can serve as a strong R-Loop initiation site. *Molecular and Cellular Biology,* 30**,** 146-159.

SAKAKIBARA, Y. & TOMIZAWA, J., I. 1974. Replication of colicin E1 plasmid DNA in cell extracts. *Proceedings for the National Academy of Sciences of the USA,* 71**,** 802-806.

SAMORA, C., P., SAKSOUK, J., GOSWAMI , P., WADE, B., O., SINGLETON, M., R., BATES, P., A., LENGRONNE, A., COSTA, A. & UHLMANN, F. 2016. Ctf4 links DNA replication with sister chromatid cohesion establishment by recruiting the Chl1 helicase to the replisome. *Molecular Cell,* 63**,** 371–384.

SARIKI, S., K., SAHU, P., K., GOLLA, U., SINGH, V., AZAD, G., K. & TOMAR, R., S. 2016. Sen1, the homolog of human Senataxin, is critical for cell survival through regulation of redox homeostasis, mitochondrial function, and the TOR pathway in *Saccharomyces cerevisiae*. *The FEBS Journal,* 283**,** 4056-4083.

SAWICKA, A. & SEISER, C. 2012. Histone H3 phosphorylation- a versatile chromatin modification for different occasions. *Biochimie,* 94**,** 2193-2201.

SAYRAC, S., VENGROVA, S., GODFREY, E., L. & DALGAARD, J., Z. 2011. Identification of a novel type of spacer element required for imprinting in fission yeast. *PLoS Genetics,* 7**,** e1001328.

SCHREIECK, A., EASTER, A., D., ETZOLD, S., WIEDERHOLD, K., LIDSCHREIBER, M., CRAMER, P. & PASSMORE, L., A. 2014. RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nature Structural and Molecular Biology,* 21**,** 175-179.

SCHWER, B. & SHUMAN, S. 2011. Deciphering the RNA Polymerase II CTD code in fission yeast. *Molecular Cell,* 43**,** 311-318.

SCHWOB, E. & NASMYTH, K. 1993. C*LB5* and C*LB6,* a new pair of B cyclins involved in DNA replication *Saccharomyces cerevisiae*. *Genes and Development,* 7**,** 1160-1175.

SEKEDAT, M., D., FENYÖ, D., ROGERS, R., S., TACKETT, A., J., AITCHISON, J., D. & CHAIT, B., T. 2010. GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Molecular Systems Biology,* 6**,** 353.

SENGUPTA, S., VAN DEURSEN, F., DE PICCOLI, G. & LABIB, K. 2013. Dpb2 integrates the leading-strand DNA polymerase into the eukaryotic replisome. *Current Biology,* 23**,** 543-552.

SHAW, N., N. & ARYA, D., P. 2008. Recognition of the unique structure of DNA:RNA hybrids. *Biochimie,* 90**,** 1026-1039.

SHIKATA, K., SASA-MASUDA, T., OKUNO, Y., WAGA, S. & SUGINO, A. 2006. The DNA polymerase activity of Pol ε holoenzyme is required for rapid and efficient chromosomal DNA replication in X*enopus e*gg extracts. *BMC Biochemistry,* 7**,** doi: 10.1186/1471-2091-7-21.

SHIMMOTO, M., MATSUMOTO, S., ODAGIRI, Y., NOGUCHI, E., RUSSELL, P. & MASAI, H. 2009. Interactions between Swi1-Swi3, Mrc1 and S phase kinase, Hsk1 may regulate cellular responses to stalled replication forks in fission yeast. *Genes to Cells,* 14**,** 669-682.

SHINKURA, R., TIAN, M., SMITH, M., CHUA, K., FUJIWARA, Y. & ALT, F., W. 2003. The influence of transcription orientation on endogenous switch function. *Nature Immunology,* 4**,** 435-441.

SIKORSKI, R., S. & HIETER, P. 1989. A system of shuttle vector and yeast host strains designed for efficient manipulation of DNA in S*accharomyces cerevisiae*. *Genetics,* 122**,** 19-27.
SIMON, A., C., SANNINO, V., COSTANZO, V. & PELLEGRINI, L. 2016. Structure of human Cdc45 and implications for CMG helicase function. *Nature Communications,* 7**,** doi: 10.1038/ncomms11638. .
SIMON, A., C., ZHOU, J., C., PERERA, R., L., VAN DEURSEN, F., EVRIN, C., IVANOVA, M., E., KILKENNY, M., L., RENAULT, L., KJAER, S., MATAK-VINKOVIĆ, D., LABIB, K., COSTA, A. & PELLEGRINI, L. 2014. A Ctf4 trimer couples the CMG helicase to DNA polymerase α in the eukaryotic replisome. *Nature,* 510**,** 293-297.
SINGH, J. & KLAR, A., J,S. 1993. DNA polymerase- a is essential for mating-type switching in fission yeast. *Nature,* 361**,** 271-273.
SKOURTI-STATHAKI, K., KAMIENIARZ-GDULA, K. & PROUDFOOT, N., J. 2014. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature,* 516**,** 436-439.
SKOURTI-STATHAKI, K., PROUDFOOT, N., J. & GROMAK, N. 2011. Human Senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Molecular Cell,* 42**,** 794-805.
SMITH, D., J. & WHITEHOUSE, I. 2012. Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature,* 483**,** 434–438.
SMITH-ROE, S., L., PATEL, S., S., SIMPSON, D., A., ZHOU, Y., C., RAO, S., IBRAHIM, J., G., KAISER-ROGERS, K., A., CORDEIRO-STONE, M. & KAUFMANN, W., K. 2011. Timeless functions independently of the Tim-Tipin complex to promote sister chromatid cohesion in normal human fibroblasts. *Cell cycle,* 10**,** 1618-1624.
SOLLIER, J., STORK, C., T., GARCÍA-RUBIO, M., L., PAULSEN, R., D., AGUILERA, A. & CIMPRICH, K., A. 2014. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Molecular Cell,* 56**,** 777-785.
SORDET, O., REDON, C., E., GUIROUILH-BARBAT, J., SMITH, S., SOLIER, S., DOUARRE, C., CONTI, C., NAKAMURA, A., J., DAS, B., B., NICOLAS, E., KOHN, K., W., BONNER, W., M. & POMMIER, Y. 2009. Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks. *EMBO Reports,* 10**,** 887-893.
STEINMETZ, E., J., CONRAD, N., K., BROW, D., A. & CORDEN, J., L. 2001. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature,* 413**,** 327-331.
STIRLING, P., C., CHAN, Y., A., MINAKER, S., W., ARISTIZABAL, M., J., BARRETT, I., SIPAHIMALANI, P., KOBOR, M., S. & HIETER, P. 2012. R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes and Development,* 26**,** 163-175.
STOLS, L., GU, M., DIECKMAN, L., RAFFEN, R., COLLART, F., R. & DONNELLY, M., I. 2002. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expression and Purification,* 25**,** 8-15.
STUCKEY, R., GARCÍA-RODRÍGUEZ, N., AGUILERA, A. & WELLINGER, R., E. 2015. Role for RNA:DNA hybrids in origin-independent replication

priming in a eukaryotic system. *Proceedings for the National Academy of Sciences of the USA,* 112**,** 5779-5784.

SUN, J., YANG, Y., WAN, K., MAO, N., YU, T., Y., LIN, Y., C., DEZWAAN, D., C., FREEMAN, B., LIN, J., J., LUE, N., F. & LEI, M. 2011. Structural basis of dimerization of yeast telomere protein Cdc13 and its interaction with the catalytic subunit of DNA polymerase a. *Cell Research,* 21**,** 258-274.

SUN, Q., CSORBA, T., SKOURTI-STATHAKI, K., PROUDFOOT, N., J. & DEAN, C. 2013. R-loop stabilization represses antisense transcription at the Arabidopsis F*LC l*ocus. *Science,* 340**,** 619-621.

SVEJSTRUP, J., Q. 2002. Mechanisms of transcription-coupled DNA repair. *Nature Reviews in Molecular and Cellular Biology,* 3**,** 21-29.

SVEJSTRUP, J., Q. 2007. Contending with transcriptional arrest during RNAPII transcript elongation. *Trends in Biochemical Sciences,* 32**,** 165-171.

SWANEY, D., L., BELTRAO, P., STARITA, L., GUO, A., RUSH, J., FIELDS, S., KROGAN, N., J. & VILLÉN, J. 2013. Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nature Methods,* 10**,** 676-682.

SZAMBOWSKA, A., TESSMER, I., KURSULA, P., USSKILAT, C., PRUS, P., POSPIECH, H. & GROSSE, F. 2014. DNA binding properties of human Cdc45 suggest a function as molecular wedge for DNA unwinding. *Nucleic Acids Research,* 42**,** 2308-2319.

SZAMBOWSKA, A., TESSMER, I., PRUS, P., SCHLOTT, B., POSPIECH, H. & GROSSE, F. 2017. Cdc45-induced loading of human RPA onto single-stranded DNA. *Nucleic Acids Research,* 45**,** 3217-3230.

SÁNCHEZ-GOROSTIAGA, A., LÓPEZ-ESTRAÑO, C., KRIMER, D., B., SCHVARTZMAN, J., B. & HERNÁNDEZ, P. 2004. Transcription termination factor Reb1p causes two replication fork barriers at its cognate sites in fission yeast ribosomal DNA *in vivo*. *Molecular and Cellular Biology,* 24**,** 398-406.

TAKAYAMA, Y., KAMIMURA, Y., OKAWA, M., MURAMATSU, S., SUGINO, A. & ARAKI, H. 2003. GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes and Development,* 17**,** 1153-1165.

TAKEUCHI, Y., HORIUCHI, T. & KOBAYASHI, T. 2003. Transcription-dependent recombination and the role of fork collision in yeast rDNA. *Genes and Development,* 17**,** 1497-1506.

TANAKA, H., KATOU, Y., YAGURA, M., SAITOH, K., ITOH, T., ARAKI, H., BANDO, M. & SHIRAHIGE, K. 2009a. Ctf4 coordinates the progression of helicase and DNA polymerase α. *Genes to Cells,* 14**,** 807-820.

TANAKA, H., KUBOTA, Y., TSUJIMURA, T., KUMANO, M., MASAI, H. & TAKISAWA, H. 2009b. Replisome progression complex links DNA replication to sister chromatid cohesion in *Xenopus* egg extracts. *Genes to Cells,* 14**,** 949-963.

TANAKA, K. & RUSSELL, P. 2001. Mrc1 channels the DNA replication arrest signal to checkpoint kinase Cds1. *Nature Cell Biology,* 3**,** 966-972.

TANAKA, S. & DIFFLEY, J., F,X. 2002. Interdependent nuclear accumulation if budding yeast Cdt1 and Mcm2-7 during G1 phase. *Nature Cell Biology,* 4**,** 198-207.

TANAKA, S., KOMEDA, Y., UMEMORI, T., KUBOTA, Y., TAKISAWA, H. & ARAKI, H. 2013. Efficient initiation of DNA replication in eukaryotes requires Dpb11/TopBP1-GINS interaction. *Molecular and Cellular Biology,* 33**,** 2614-2622.

TANAKA, S., UMEMORI, T., HIRAI, K., MURAMATSU, S., KAMIMURA, Y. & ARAKI, H. 2007. CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature,* 445**,** 328-332.

TENNYSON, C., N., KLAMUT, H., J. & WORTON, R., G. 1995. The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nature Genetics,* 9**,** 184-190.

TERCERO, J., A. & DIFFLEY, J., F,X. 2001. Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint. *Nature,* 412**,** 553-557.

TEWARI, R., BAILES, E., BUNTING, K., A. & COATES, J., C. 2010. Armadillo-repeat protein functions: questions for little creatures. *Trends in Cell Biology,* 20**,** 470-481.

THEIS, J., F. & NEWLON, C., S. 1997. The *ARS309* chromosomal replicator of S*accharomyces cerevisiae* depends on an exceptional *ARS* consensus sequence. *Proceedings for the National Academy of Sciences of the USA,* 94**,** 10786-10791.

THON, G., BJERLING, K., P. & NIELSEN, I., S. 1999. Localization and properties of a silencing element near the mat3-M mating-type cassette of S*chizosaccharomyces pombe*. *Genetics,* 151**,** 945-963.

THON, G., COHEN, A. & KLAR, A., J,S. 1994. Three additional linkage groups that repress transcription and meiotic recombination in the mating-type region of S*chizosaccharomyces pombe*. *Genetics,* 138**,** 29-38.

THON, G. & KLAR, A., J,S. 1993. Directionality of fission yeast mating-type interconversion is controlled by the location of donor loci. *Genetics,* 134**,** 1045-1054.

TOMIZAWA, J., I. 1975. Two distinct mechanisms of synthesis of DNA fragments on colicon E1 plasmid DNA. *Nature,* 257**,** 253-254.

TURNER, R., M,JR., GRINDLEY, N., D. & JOYCE, C., M. 2003. Interaction of DNA polymerase I (Klenow fragment) with the single-stranded template beyond the site of synthesis. *Biochemistry,* 42**,** 2373-2385.

TYE, B., K. 1999. Minichromosome maintenance as a genetic assay for defects in DNA replication. *Methods,* 18**,** 329-334.

TYERS, M., TOKIWA, G. & FUTCHER, B. 1993. Comparison of the *Saccharomyces cerevisiae* G1 cyclins: Cln3 may be an upstream activator of Cln1, Cln2 and other cyclins. *The EMBO Journal,* 12**,** 1955-1968.

URSIC, D., CHINCHILLA, K., FINKEL, J., S. & CULBERTSON, M., R. 2004. Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic Acids Research,* 32**,** 2441-2452.

USDIN, K. & WOODFORD, K., J. 1995. CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis i*n vitro*. *Nucleic Acids Research,* 23**,** 4202-4209.

VAN DEURSEN, F., SENGUPTA, S., DE PICCOLI, G., SANCHEZ-DIAZ, A. & LABIB, K. 2012. Mcm10 associates with the loaded DNA helicase at

replication origins and defines a novel step in its activation. *The EMBO Journal,* 31**,** 2195-2206.

VAN GOOL, A., J ., VERHAGE, R., SWAGEMAKERS, S., M., VAN DE PUTTE, P., BROUWER, J., TROELSTRA, C., BOOTSMA, D. & HOEIJMAKERS, J., H . 1994. R*AD26,* the functional S*. cerevisiae* homolog of the Cockayne syndrome B gene ERCC6. *The EMBO Journal,* 13**,** 5361-5369.

VANNINI, A. & CRAMER, P. 2012. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Molecular Cell,* 45**,** 439-446.

VASILJEVA, L. & BURATOWSKI, S. 2006. Nrd1 interacts with the nuclear exosome for 3' processing of RNA Polymerase II transcripts. *Molecular Cell,* 21**,** 239-248.

VASILJEVA, L., KIM, M., MUTSCHLER, H., BURATOWSKI, S. & MEINHART, A. 2008. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nature Structural and Molecular Biology,* 15**,** 795-804.

VENGROVA, S. & DALGAARD, J., Z. 2004. RNase-sensitive DNA modification(s) initiates S*. pombe* mating-type switching. *Genes and Development,* 18**,** 794-804.

VENGROVA, S. & DALGAARD, J., Z. 2005. The S*chizosaccharomyces pombe* imprint--nick or ribonucleotide(s)? *Current Biology,* 15**,** R326-R327.

VENGROVA, S. & DALGAARD, J., Z. 2006. The w*ild-type Schizosaccharomyces pombe* m*at1* imprint consists of two ribonucleotides. *The EMBO Reports,* 7**,** 59-65.

VILLA, F., SIMON , A., C., BAZAN, M., A,O., KILKENNY, M., L., WIRTHENSOHN, D., WIGHTMAN, M., MATAK-VINKOVÍC, D., PELLEGRINI, L. & LABIB, K. 2016. Ctf4 is a hub in the eukaryotic replisome that links multiple CIP-box proteins to the CMG helicase. *Molecular Cell,* 63**,** 385-396.

VIVIAN, J., P., PORTER, C., J., WILCE, J., A. & WILCE, M., C. 2007. An asymmetric structure of the B*acillus subtilis* replication terminator protein in complex with DNA. *Journal of Molecular Biology,* 370**,** 481-491.

WAGA, S., BAUER, G. & STILLMAN, B. 1994. Reconstitution of complete SV40 DNA replication with purified replication factors. *The Journal of Biological Chemistry,* 269**,** 10923-10934.

WAKE, R., G. 1997. Replication fork arrest and termination of chromosome replication in B*acillus subtilis*. *FEMS Microbiology Letters,* 153**,** 247-254.

WANG, R., YANG, B. & ZHANG, D. 2011. Activation of interferon signaling pathways in spinal cord astrocytes from an ALS mouse model. *Glia,* 59**,** 946–958

WARBRICK, E. 1998. PCNA binding through a conserved motif. *Bioessays,* 20**,** 195-199.

WATASE, G., TAKISAWA, H. & KANEMAKI, M., T. 2012. Mcm10 plays a role in functioning of the eukaryotic replicative DNA helicase, Cdc45-Mcm-GINS. *Current Biology,* 22**,** 343-349.

WATSON, A., T., GARCIA, V., BONE, N., CARR, A., M. & ARMSTRONG, J. 2008. Gene tagging and gene replacement using recombinase-mediated cassette exchange in S*chizosaccharomyces pombe*. *Gene,* 407**,** 63-74.

WATT, D., L., JOHANSSON, E., BURGERS, P., M. & KUNKEL, T., A. 2011. Replication of ribonucleotide-containing DNA templates by yeast replicative polymerases. *DNA Repair,* 10**,** 897-902.

WEAVER, D., T. & DEPAMPHILIS, M., L. 1982. Specific sequences in native DNA that arrest synthesis by DNA polymerase α. *The Journal of Biological Chemistry,* 257**,** 2075-2086.

WEBB, S., HECTOR, R., D., KUDLA, G. & GRANNEMAN, S. 2014. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biology,* 15**,** R8.

WEINERT, T., A., KISER, G., L. & HARTWELL, L., H. 1994. Mitotic checkpoint genes in budding yeast and the dependence of mitosis on DNA replication and repair. *Genes and Development,* 8**,** 652-665.

WEINREICH, M. & STILLMAN, B. 1999. Cdc7p-Dbf4p kinase binds to chromatin during S phase and is regulated by both the APC and the RAD53 checkpoint pathway. *The EMBO Journal,* 18**,** 5334–5346.

WEITZMANN, M., N., WOODFORD, K., J. & USDIN, K. 1997. DNA secondary structures and the evolution of hypervariable tandem arrays. *The Journal of Biological Chemistry,* 272**,** 9517-9523.

WELLINGER, R., E., PRADO, F. & AGUILERA, A. 2006. Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Molecular and Cellular Biology,* 26**,** 3327-3334.

WEST, S., GROMAK, N. & PROUDFOOT, N., J. 2004. Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature,* 432**,** 522-525.

WESTOVER, K., D., BUSHNELL, D., A. & KORNBERG, R., D. 2004. Structural basis of transcription : nucleotide selection by rotation in the RNA Polymerase II active center. *Cell,* 119**,** 481-489.

WHITELAW, E. & PROUDFOOT, N. 1986. α-Thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human α2 globin gene. *The EMBO Journal,* 5**,** 2915-2922.

WILCE, J., A., VIVIAN, J., P., HASTINGS, A., F., OTTING, G., FOLMER, R., H., DUGGIN, I., G., WAKE, R., G. & WILCE, M., C. 2001. Structure of the RTP-DNA complex and the mechanism of polar replication fork arrest. *Nature Structural Biology,* 8**,** 206-210.

WILD, T. & CRAMER, P. 2012. Biogenesis of multisubunit RNA polymerases. *Trends in Biochemical Sciences,* 37**,** 99-105.

WILLIAMS, D., R. & MCINTOSH, J., R . 2002. m*cl1+*, the Schizosaccharomyces pombe homologue of C*TF4,* is important for chromosome replication, cohesion, and segregation. *Eukaryotic Cell,* 1**,** 758-773.

WILLIAMS, D., R. & MCINTOSH, J., R. 2005. Mcl1p is a polymerase alpha replication accessory factor important for S-phase DNA damage survival. *Eukaryotic Cell,* 4**,** 166-177.

WIMBERLY, H., SHEE, C., THORNTON, P., C., SIVARAMAKRISHNAN, P., ROSENBERG, S., M. & HASTINGS, P., J. 2013. R-loops and nicks initiate DNA breakage and genome instability in non-growing E*scherichia coli*. *Nature Communications,* 4**,** 2115.

WING, R., DREW, H., TAKANO, T., BROKA, C., TANAKA, S., ITAKURA, K. & DICKERSON, R., E. 1980. Crystal structure analysis of a complete turn of B-DNA. *Nature,* 287**,** 755-758.

WOODFORD, K., J., HOWELL, R., M. & USDIN, K. 1994. A novel K^+-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. *The Journal of Biological Chemistry,* 269**,** 27029-27035.

XU, B. & CLAYTON, D., A. 1996. RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers. *The EMBO Journal,* 15**,** 3135-3143.

XU, H., BOONE, C. & BROWN, G., W. 2007. Genetic dissection of parallel sister-chromatid cohesion pathways. *Genetics,* 176**,** 1417-1429.

XU, X., BLACKWELL, S., LIN, A., LI, F., QIN, Z. & XIAO, W. 2015. Error-free DNA-damage tolerance in S*accharomyces cerevisiae*. *Mutation Research/ Reviews in Mutation Research,* 764**,** 43–50.

YACHDAV, G., KLOPPMANN, E., KAJAN, L., HECHT, M., GOLDBERG, T., HAMP, T., HÖNIGSCHMID, P., SCHAFFERHANS, A., ROOS, M., BERNHOFER, M., RICHTER, L., ASHKENAZY, H., PUNTA, M., SCHLESSINGER, A., BROMBERG, Y., SCHNEIDER, R., VRIEND, G., SANDER, C., BEN-TAL, N. & ROST, B. 2014. PredictProtein--an open resource for online prediction of protein structural and functional features. *Nucleic Acids Research* 42**,** W337-W343.

YAMADA-INAGAWA, T., KLAR, A., J,S. & DALGAARD, J., Z. 2007. S*chizosaccharomyces pombe* switches mating type by synthesis-dependent strand annealing mechanism. *Genetics,* 177**,** 255-265.

YANG, X., GREJAN, J., LINDNER, K., YOUNG, H. & KEARSEY, S., E. 2005. Nuclear distribution and chromatin association of DNA polymerase α-primase is affected by TEV protease cleavage of Cdc23 (Mcm10) in fission yeast. *BMC Molecular Biology,* 6**,** 13.

YAO, N., TURNER, J., KELMAN, Z., STUKENBERG, P., T., DEAN, F., SHECHTER, D., PAN, Z., Q., HURWITZ, J. & O'DONNELL, M. 1996. Clamp loading, unloading and intrinsic stability of the PCNA, beta and gp45 sliding clamps of human, *E. coli* and T4 replicases. *Genes to Cell,* 1**,** 101-113.

YEELES, J., T,P. & MARIANS, K., J. 2013. Dynamics of leading-strand lesion skipping by the replisome. *Molecular Cell,* 52**,** 855-865.

YEELES, J., T., JANSKA, A., EARLY, A. & DIFFLEY, J., F. 2017. How the eukaryotic replisome achieves rapid and efficient DNA replication. *Molecular Cell,* 65**,** 105-116.

YEO, A., J ., BECHERE, L. O., J., LUFF, J., E ., CULLEN, J., .K ., WONGSURAWAT, T., JENJAROENPUN, P., KUZNETSOV, V., A ., MCKINNON, P., J . & LAVIN, M., F . 2014. R-loops in proliferating cells but not in the brain: implications for AOA2 and other autosomal recessive ataxias. *PLoS ONE,* 9**,** e90219.

YU, K. & LIEBER, M., R. 2003. Nucleic acid structures and enzymes in the immunoglobulin class switch recombination mechanism. *DNA Repair,* 2**,** 1163-1174.

YU, S., L., KANG, M., S., KIM, H., Y., GOROSPE, C., M., KIM, T., S. & LEE, S., K. 2014. The PCNA binding domain of Rad2p plays a role in mutagenesis by modulating the cell cycle in response to DNA damage. *DNA Repair,* 16**,** 1-10.

YURIEVA, O. & O'DONNELL, M. 2016. Reconstitution of a eukaryotic replisome reveals the mechanism of asymmetric distribution of DNA polymerases. *Nucleus,* 7**,** 360-368.

YURYEV, A., PATTURAJAN, M., LITINGTUNG, Y., JOSHI, R., V., GENTILE, C., GEBARA, M. & CORDEN, J., L. 1996. The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine rich proteins. *Proceedings for the National Academy of Sciences of the USA,* 93**,** 6975-6980.

YÜCE, Ö. & WEST, S., C. 2013. Senataxin, defective in the neurodegenerative disorder Ataxia with Oculomotor Apraxia 2, lies at the interface of transcription and the DNA damage response. *Molecular and Cellular Biology,* 33**,** 406 -417.

ZECH, J., GODFREY, E., L., MASAI, H., HARTSUIKER, E. & DALGAARD, J., Z. 2015. The DNA-binding domain of S*. pombe* Mrc1 (Claspin) acts to enhance stalling at replication barriers. *PLoS ONE,* 10**,** e0132595.

ZEGERMAN, P. & DIFFLEY, J., F. 2010. Checkpoint-dependent inhibition of DNA replication initiation by Sld3 and Dbf4 phosphorylation. *Nature,* 467**,** 474-478.

ZHAO, H., TANAKA, K., NOGOCHI, E., NOGOCHI, C. & RUSSELL, P. 2003. Replication checkpoint protein Mrc1 is regulated by Rad3 and Tel1 in fission yeast. *Molecular and Cellular Biology,* 23**,** 8395-8403.

ZHOU, Y. & WANG, T., S,F. 2004. A coordinated temporal interplay of nucleosome reorganization factor, sister chromatin cohesion factor and DNA Polymerase α facilitates DNA replication. *Molecular and Cellular Biology,* 24**,** 9568-9579.

ZHU, W., UKOMADU, C., JHA, S., SENGA, T., DHAR, S., K., WOHLSCHLEGEL, J., A., NUTT, L., K., KORNBLUTH, S. & DUTTA, A. 2007. Mcm10 and And-1/CTF4 recruit DNA polymerase alpha to chromatin for initiation of DNA replication. *Genes and Development,* 21**,** 2288-2299.